Genome Wide Gene Expression Analysis of Two ENU Mouse Models of Major Mental Illness

Sarah Mills Brown

PhD
The University of Edinburgh
2011

Declaration

I declare that this thesis has been composed by myself, that the work described in this thesis is my own, except where otherwise stated, and that the work described in this thesis has not been submitted for any other degree or personal qualification.

Sarah Mills Brown

Acknowledgements

So many people have supported me, in one way or another, over the course of this PhD that I am afraid I cannot possibly do them all justice here. I am forever grateful and indebted to the friends, family and colleagues who have in many cases gone beyond the call of duty to help me get to where I am today. I could not have done it without you and I thank you all.

To my supervisors, Kathy and Pippa, for your support, supervision and patience. To David for your encouragement and support, and for pushing me to believe in myself. To Helen and Susan, for answering my unending questions and supporting and training me in the early stages. To Steve and Stewart for their computing expertise. I am sure you will be breathing a sigh of relief that I am finished! To John Roder and Steve Clapcote who welcomed me into their lab in Toronto while I was collecting samples and who taught me everything I know about managing mouse colonies! To Angie, Allison and Lee from the WTCRF for helping set up and run the microarray study. To William from the WTCRF for training and assistance for the real-time PCR, and ensuring there was always a space in the booking form for me to complete my experiments on time. To Tom Freeman for microarray statistical assistance, and Andrea Christoforou for all statistical assistance! To Sharon, Rosemary, Helen and Susan (again!) for taking care of absolutely everything else so I could concentrate on my work! To Becky, who was a constant sounding board and friend, and who on many occasions helped me see the good side when things weren't going to plan.

To my wonderful friends who have provided both encouragement and support when I have sometimes been less than easy to be around, and to my partner Gordon, who must have the patience of a saint to put up with my stressful flapping! To Jake and Taff, for keeping me grounded and helping me to take time out to relax when things got a little hectic.

Finally, I would not have been able to do this without the support of my amazing family. You will possibly never know how much the seemingly pointless phone calls, emails and photos from home have meant to me. This thesis is dedicated to you.

Thank you all.

Abstract

Major mental illness is now recognised as one of the leading causes of adult morbidity. Of the adult onset psychiatric disorders, the functional psychoses (schizophrenia, bipolar disorder and recurrent major depression) are the most severe and most common in the general population. Evidence suggests that certain genetic factors influence an individual's susceptibility to developing these disorders when combined with appropriate social and environmental conditions. Several good candidate genes have been identified. Of relevance to this study is Disrupted in Schizophrenia 1 (DISC1) which was identified in a large Scottish family that carried a balanced translocation (t1:11) and had a history of major mental illness. In 2008, two ENU mutant mouse models with missense mutations in exon 2 of Disc1 were characterised and found to have behavioural and neuroanatomical phenotypes consistent with schizophrenia and major depression. The primary aim of this thesis is to further analyse these mouse models by performing whole genome gene expression studies and secondary protein analysis to identify genes involved in the aetiology of schizophrenia and major depression.

My initial analysis used Illumina BeadChip microarray technology to identify 368 genes that were differentially expressed in ENU mutant animals under different biological conditions, compared to appropriate control animals. Nine biological groups were compared including one embryonic group at E13, and three groups treated with appropriate anti-psychotic or anti-depressant drugs. Of the 368 genes identified as differentially expressed, 46 were chosen for validation by qRT-PCR based on fold-change, p-value, functional significance, overenrichment of GO terms, pathway analysis and previous implications in major mental illness. NRXN1, NRXN3 and CDH11 were found to be significantly up-regulated in the schizophrenia mouse model with EGR4 significantly down-regulated compared to C57BL/6J wild-type controls. These findings were also replicated in an independent sample using wild-type littermates. The mental retardation gene PAK3 was up-regulated in the schizophrenia mouse model and expression levels were corrected to a level not significantly different to wild-type, when treated with the PDE4 inhibitor Rolipram. Semi-quantitative western blotting also confirmed the disregulation of EGR4 and PAK3 at the protein level in these animals. RNA expression profiles were also characterised for each of the genes above, and DISC1, through development.

In summary this thesis describes the striking disregulation of four prominent genetic candidates of major mental illness in an independent animal model. A first functional link between DISC1 and NRXN1 is described suggesting, for the first time, a DISC1-dependant mechanism for regulating neurexin gene expression.

Table of Contents

Declaration……ii
Acknowledgements……...iii
Abstract……..……...iv
Table of Contents……..vi
List of Figures…….xiv
List of Tables…….xvii
Abbreviations.....xix

1. Introduction

1.1 Major Mental Illness Overview……2
1.2 A genetic basis for major mental illness……..6
1.2.1 Linkage Analysis……...8
1.2.2 Association Studies……...8
1.2.3 Copy Number Variation……..11
1.2.4 Genome wide gene expression analysis……..12
1.2.5 Cytogenetics.....14
1.3 Identification of Disrupted in Schizophrenia 1……..14
1.4 The biology of DISC1 and its interactions…….18
1.5 Current Hypotheses of Major Mental Illness…….22
1.5.1 The neurotransmitter hypothesis.....22
1.5.2 The developmental hypothesis…...…….23
1.5.3 The cAMP hypothesis….....……24
1.6 Genome-wide expression and pathway analysis……..…….25
1.7 The schizophrenia mouse……...26
1.8 Thesis Aims……35

2. Materials and Methods

2.1 Collection of Tissue Samples…….37

2.1.1 Use of inbred strains……37

2.1.2 Extraction of Genomic DNA from adult tissue……...38

2.1.3 Extraction of Genomic DNA from embryonic tissue……..38

2.1.4 Genotyping by sequencing reaction……39

2.1.4.1 Polymerase Chain Reaction…….39

2.1.4.2 Exosapit Clean-up……39

2.1.4.3 Sequencing reaction…….40

2.1.4.4 Ethanol/EDTA Precipitation……40

2.1.5 Genotyping using the Transnetyx sequencing service……40

2.1.6 Genotyping Q31L het x hom pups……..41

2.1.7 Dissection of embryos and pups……..41

2.1.8 Dissection of adult brain samples……41

2.2 Sample Preparation……42

2.2.1 Extraction of total RNA from tissue samples……..42

2.2.2 PARIS procedure…….43

2.2.3 Post extraction DNase Treatment of RNA……..43

2.2.4 On column DNase Treatment of RNA……44

2.2.5 Assessing RNA quality using the Agilent Bioanalyser……...44

2.2.6 Vacuum Drying RNA samples……46

2.2.7 Synthesis of cDNA……..47

2.2.8 Synthesis of cDNA from low quantity RNA……..47

2.2.9 Determining cDNA concentration using the Quant-iT™ PicoGreen® dsDNA kit ……48

2.2.10 Determining cDNA or RNA quantity on the NanoDrop…...…49

2.3 PCR……50

2.3.1 PCR Reagents……..50

2.3.2 Design and Synthesis of Oligonucleotide Primer pairs……...50

2.3.3 RT-PCR Cycling…….51

2.3.4 Visualisation of PCR Products……51

2.4 Protein Preparations……...52

2.4.1 Determining Protein Sample Concentrations……..52

2.5 Pre-Pulse Inhibition……53

2.5.1 Calibration……...53

2.5.2 Drug Administration……...54

2.5.3 Behavioural Testing……54

2.5.4 Analysis of Behavioural Testing Results……55

2.6 Microarray Analysis of ENU mutant gene expression……...55

2.6.1 cRNA amplification and Biotinylation……...55

2.6.2 Quantification and Quality control of cRNA……..56

2.6.3 Preparation of Pooled cRNA……...56

2.6.4 Hybridisation and scanning of Illumina BeadChips at the Wellcome Trust Clinical Research Facility……..57

2.6.5 Data Analysis……...…58

2.6.6 Microarray Quality Control Measures……58

2.6.6.1 Illumina direct hybridisation assay controls…….58

2.6.6.2 Technical Replicates and Between Chip Controls…...58

2.6.6.3 Cluster analysis for detection of experimental Artefacts……...59

2.6.6.4 Power Calculations……...59

2.6.6.5 Pre-processing……..59

2.7 Testing For Differential Expression……..61

2.7.1 Differential expression analysis……..61

2.7.2 Functional Analysis (Including IPA Pathway analysis)……..61

2.7.3 GOTree Analysis of differentially expressed genes……62

2.7.4 Testing for overlaps and gene expression corrections from drug treatments……62

2.7.5 Comparing data with previously published work……...63

2.8 Follow up analysis of the microarray study……...63

2.8.1 Identifying the genes to take forward for validation……...63

2.8.2 Preparation of Samples for two phase follow up……64

2.8.3 Identification of suitable probes for Taqman real time PCR…..64

2.8.4 Verification of genes by ABI Taqman analysis……..65

2.8.5 Statistical Analysis of ABI Taqman output……66

2.9 Protein analysis and Antibody Staining…….67

2.9.1 Identification of suitable antibodies……67

2.9.2 Western Blotting…….68

2.9.3 Quantification of Westerns using Image J……..69

2.9.4 Culture of Primary Neurons……69

2.9.5 ICC of Primary Neurons…….70

2.10 Buffers……..71

2.10.1 Western Blotting……...71

2.10.2 Primary Neuron Production……..72

3. Sample collection and processing of a genome wide microarray analysis of the Disc1 ENU mouse mutants

3.1 Introduction……...…….75

3.1.1 Background and motivation……75

3.2 Partial Replication of Clapcote *et al's* Behavioural Analysis of the Disc1 ENU mouse mutants……80

3.2.1 Effect of Gender on PPI in non-drug treated animals…….80

3.2.2 Effect of Genotype on PrePulse Inhibition and Startle Response in Mice……..80

3.2.3 Effect of Drug treatment on PrePulse Inhibition and Acoustic Startle Response in Mice……..82

3.3 Whole Genome Gene Expression Study Experimental Design…….83

3.3.1 Sample size and power……83

3.3.2 Sample Layout and Technical Replication……..86

3.4 Sample Preparation and Quality…….87

3.4.1 Integrity of total RNA…….88

3.4.2 Integrity of cRNA……91

3.5 Data Generation using the Illumina BeadArray Platform……..92

3.6 Quality Control……..95

3.6.1 Illumina BeadChip Built in Controls……..95

3.6.2 Technical Replicates……..…101
3.6.3 Quality Control using Cluster Analysis……103

3.7 Data Pre-processing…….106

3.7.1 Variance Stabilisation……...106

3.7.2 Normalisation……...107

3.7.3 Data Filtering……109

3.8 Conclusions……..112

4. Differential Expression Analysis of the Disc1 ENU Mutant Mouse Microarray
4.1 Introduction……..117
4.2 Differentially Expressed Genes……117
4.2.2 Differentially expressed genes: primary
annotation and gene ontology……119
4.2.3 Differential Expression between the L100P
adult ENU mutant and the C57BL/6J adult controls…….121
4.2.4 Differential Expression between the Q31L
adult ENU mutant and the C57BL/6J adult controls…….136
4.2.5 Differential Expression between the Embryonic
ENU mutant and the C57BL/6J adult controls……..136
4.2.6 The effect of Drug Treatment on gene expression
in adult ENU mutant mice……..138
4.3 Selection of Genes of Interest for Validation by qRT PCR…….145
4.4 Identification of L100P outlier…….147
4.4.1 Reanalysis of array dataset……148
4.4.2 New Gene Ontology, Pathway analysis and
Comparisons with previous study……..157
4.5 New list of genes for qRT PCR……....160
4.6 Conclusions…...……..162

5. Quantitative real-time PCR analysis of candidate differentially expressed genes in the L100P and Q31L mutant mice

5.1 Introduction.....169

5.2 Validating the Disc1 microarray result.....170

5.2.1 Initial Validations and Outlier removal of the L100P drug naïve adult microarray.....170

5.2.2 Validation of the Q31L drug naïve adult Microarray.....173

5.2.3 Validation of the L100P and Q31L embryonic Microarray.....174

5.2.4 Validation of the L100P and Q31L drug treatment Microarray.....174

5.3 Conclusions.....177

6. Establishing a developmental profile for the genes of interest in L100P mutant mice and wild type controls.....183

6.1 The Disc1 developmental profile.....183

6.2 Establishing developmental profiles for the genes of interest.....186

6.2.1 The Disc1 developmental profile on pooled samples.....186

6.2.2 Developmental Profile for EGR4.....189

6.2.3 Developmental Profile for NRXN1 and NRXN3.....190

6.2.4 Developmental Profile of PAK3.....191

6.2.5 Developmental Profile of Cdh11.....193

6.2.6 Developmental Profile of Sort1.....194

6.3 Conclusions from the developmental profiles.....195

7. Preliminary protein analysis of validated genes in L100P drug naïve (saline) adult mouse samples.....199

7.1 Protein analysis of Egr4.....199

7.2 Protein analysis of Pak3.....201

7.3 Protein analysis of NRXN1.....202

7.4 Protein analysis of Nrxn3.....204

7.5 Protein analysis of Sort1.....206

7.6 Conclusions.....207

8. Discussion

8.1 Summary of findings……211

8.1.1 Chapter 3……..211

8.1.2 Chapter 4……..212

8.1.3 Chapter 5……..216

8.1.4 Chapter 6……..219

8.1.5 Chapter 7……..221

8.1.6 Conclusions……..223

8.2 Caveats…….223

8.3 Relevance of this study to the field……..227

8.4 Differential networking and other future work……229

8.5 Final Remarks……..231

Appendix……232

References……..233

List of Figures

Figure 1.1: The Disc1 protein……..32

Figure 2.1: Examples of Agilent 2100 Bioanalyser output Electropherographs……..46

Figure 2.2: Picogreen standard curve for assessing cDNA concentration……...49

Figure 2.3: Image of a PrePulse inhibition chamber……53

Figure 3.1: Image depicting the effective reduction in an animals startle response when a strong stimulus is preceded by a weaker prestimulus…….78

Figure 3.2: Total PrePulse Inhibition and Acoustic Startle Response in non-drug treated adult mice from each line…….81

Figure 3.3: Power Curves…….85

Figure 3.4: Layout of the seven Illumina mouse ref-6 version 2 BeadChips……...87

Figure 3.5: Examples of Agilent 2100 Bioanalyser output Electropherographs……..89

Figure 3.6: Agilent Bioanalyser output of cRNA Integrity……..92

Figure 3.7: Signal Intensities for Illumina built-in Quality control measures……..96

Figure 3.8: Signal Intensities for Illumina built-in Quality control measures by BeadChip……..…99

Figure 3.9: Signal Intensities for Illumina built-in Quality control measures by BeadChip post chip5 repeat……...100

Figure 3.10: Pairwise Comparisons of raw data for C57BL6 F1 pool samples…….102

Figure 3.11: Pairwise comparisons of raw data for C57BL6 F2 pool samples……..103

Figure 3.12: Heirarchical cluster dendrogram of all samples based on the raw dataset……..105

Figure 3.13: Heirarchical cluster dendrogram of all samples based on the stabilized and normalized dataset……...108

Figure 3.14: Heirarchical cluster dendrogram of all samples based on the filtered, transformed and stabilized dataset……..111

Figure 3.15: Venn diagrams of probes surviving the 1.3 fold-change filter in six comparisons…….112

Figure 4.1 Venn diagram of significant probes in the genotype comparisons……...119

Figure 4.2: Bar graph of GOTree catagories for L100Padult vs C57BL6adult…….132

Figure 4.3: Ingenuity Pathway Analysis networks for L100Padult disregulated Genes……..……134

Figure 4.5: Ingenuity Pathway Analysis of Embryonic disregulated gene lists…….138

Figure 4.6: Ingenuity Pathway Analysis of Drug treated groups……...144

Figure 4.7: PrePulse Inhibition data for L100PSaline and C57BL/6J mice……...148

Figure 4.8: Venn diagram of significant probes in the genotype comparisons……..156

Figure 4.9: Ingenuity Pathway Analysis……158

Figure 4.10: Canonical pathways for IPA analysis on new disregulated gene list…159

Figure 6.1: Developmental Profile of Disc1 in the mouse brain……184

Figure 6.2: Comparing Disc1 expression in pooled and unpooled developmental samples……..187

Figure 6.3: Developmental profile of Disc1 in the L100P mutant mouse…….188

Figure 6.4: Developmental profile of *Egr4* in cDNA pools from the L100P mutant mouse…….189

Figure 6.5: Developmental profile of Nrxn1 and Nrxn3 in cDNA pools from the L100P mutant mouse…….190

Figure 6.6: Developmental profile of PAK3 in cDNA pools from the L100P mutant mouse……..192

Figure 6.7: Developmental profile of Cdh11 in cDNA pools from the L100P mutant mouse……..…193

Figure 6.8: Developmental profile of Sort1 in cDNA pools from the L100P mutant mouse……..194

Figure 6.9 Composite of the developmental profiles compared to Disc1.....196

Figure 7.1: Western blotting and ICC for EGR4 in L100P ENU mutant mice……..200

Figure 7.2: Western blotting and ICC for Pak3 in L100P ENU mutant mice……201

Figure 7.3: Western blotting and ICC for Nrxn1 in L100P ENU mutant mice…….203

Figure 7.4: Western blotting and ICC for NRXN3 in L100P ENU mutant mice…..205

Figure 7.5: Western blotting and ICC for Sort1 in L100P ENU mutant mice……...206

List of Tables

Table 1.1: Phenotypes of *DISC1/Disc1* mouse models……30

Table 1.2: Common behavioural traits and the effects of the ENU missense mutation on phenotype when compared to the C57B6/J parent line …….33

Table 1.3: The effects of drug treatment on behavioural rescue in Q31L and L100P missense mutant mouse lines……..34

Table 2.1: BeadChip layout with key……..56

Table 3.1: Behavioural Phenotypes of the L100P and Q31L Disc1 mutant mice…..79

Table 3.2: Mean results on PPI at all 3 levels tested, plus total……..82

Table 3.3: Mean acoustic startle response and average PPI over all ranges tested….82

Table 3.4 Summary of Adult Mouse Groups used in the genome-wide expression analysis……...83

Table 3.5 Summary of developmental mouse groups used in the genome wide expression analysis……84

Table 3.6: Mouse samples used in the microarray analysis…….90

Table 3.7: Summary of raw expression data……94

Table 4.1 Summary of probes differentially expressed by genotype of drug treatment……118

Table 4.3: Complete list of disregulated genes in the L100P adult group……122

Table 4.2: Overlap with previous studies……..131

Table 4.4: Disregulated genes in the Q31L adult group compared to C57BL/6J controls……136

Table 4.5: Disregulated genes in the Embryonic sample groups……..137

Table 4.6.1: Disregulated genes in Clozapine treated L100P adults…….139

Table 4.6.2: Disregulated genes in Rolipram treated L100P adults……..140

Table 4.6.3: Disregulated genes in Bupropion treated Q31L adults……..141

Table 4.7: Genes for validation by qRT PCR……....146

Table 4.8 Summary of probes differentially expressed by genotype

of drug treatment……...149

Table 4.9: New list of disregulated genes post outlier removal……...150

Table 4.10: Table of biological processes over-enriched in the

L100PSaline group……157

Table 4.11: Overlap with previous studies……160

Table 4.12: List of genes to be analysed by qRT PCR…….…….161

Table 5.1: Table of genes differentially expressed in the L100P drug

naïve adult that were tested by qRT-PCR……..172

Table 5.2: Table of genes differentially expressed in the Q31L drug

naïve adult that were tested by qRT-PCR. ……173

Table 5.3: Table of genes differentially expressed in the L100P and

Q31L embryo that were tested by qRT-PCR……..…..174

Table 5.4: Table of genes differentially expressed in the L100P and

Q31L drug treated adults that were tested by qRT-PCR……..175

Table 5.5: Table of genes differentially expressed in the C57BL/6J

drug treated adults that were tested by qRT-PCR……....176

Abbreviations

ANOVA	Analysis of Variance
cDNA	Complimentary DNA
CNS	Central nervous system
CNV	Copy Number Variant
ENU	N-ethyl-N-nitrosourea
ER	Endoplasmic reticulum
FDA	Food and drugs agency
FDR	False discovery rate
fMRI	functional magnetic resonance imaging
GO	Gene ontology
HapMap	Haplotype map
ICC	Immunocytochemistry
IQR	Interquartile range
LOD	Logarithm of the odds
MAANOVA	Microarray analysis of variance
mRNA	Messenger RNA
PPI	Prepulse inhibition
qRT-PCR	Quantitative real time polymerase chain reaction
RIN	RNA integrity number
RNAi	RNA interference
SNP	Single nucleotide polymorphism
Vst	Variance stabilising transformation

Chapter 1
Introduction

1.1 Major Mental Illness Overview

Major mental illness is now recognised as one of the leading causes of adult morbidity, affecting populations throughout the world [1] with little demographic variability. Of the adult onset psychiatric disorders, the functional psychoses (schizophrenia, bipolar disorder and recurrent major depression) are the most severe and the most common in the general population. Functional psychoses are serious disorders with, as yet, no known organic cause [2] often rendering the sufferer unable to maintain normal day to day activities. They encompass disorders of thought, perception, often leading to delusion and hallucination, mood and behaviour. Evidence suggests that genetic factors influence an individual's susceptibility to developing these disorders [3] when combined with appropriate social or environmental conditions. Over the last century the functional psychoses have classically been divided into two diagnostic groups; schizophrenia and affective disorders (recurrent major depression and bipolar disorder). More recently, it has been suggested that the disorders may not be as distinct as was once thought but form a continuum with recurrent major depression at one end and schizophrenia at the other [4]. This view is supported by epidemiological evidence that shows genetic sharing of risk between schizophrenia and bipolar disorder [5], increased risk of both schizophrenia and bipolar disorder in first degree probands with schizophrenia or bipolar disorder [6], recent genome wide association studies [7] and indeed earlier work from our own laboratory [8].

Schizophrenia is the most common form of psychotic illness with a lifetime morbid risk of around 1% (more if the many schizophrenia spectrum disorders are also considered) and roughly equal proportions of males and females affected. The illness is characterised by behavioural and cognitive symptoms, which can be categorised as either positive or negative (referring to an increase or absence in normal functioning). Positive symptoms include heightened sensory perception, often leading to delusion and hallucination, disorganisation, rapidity of speech, uncoordinated gait and altered cognitive functioning. Negative symptoms include social withdrawal, lack of care over personal appearance or wellbeing and the inability to act to achieve simple goals [9]. Diagnosis of schizophrenia is based on patient interview with presentation of at least two symptoms over a period of

2-6 months affecting ability to work, socialise or provide basic self care. Many schizophrenia patients are permanently or temporarily hospitalised as they may pose a danger to themselves, and current statistics show 10% of schizophrenia patients commit suicide and many more attempt to [10]. The need for long term social and medical care of schizophrenia patients costs the health service around £1 billion per year in England and Wales alone [11]. Onset of psychosis is normally in adolescence to mid-twenties after the end of puberty, with male patients being admitted on average 4-5 years earlier than female patients [12]. Some studies have concluded that the symptomology of schizophrenia differs between the sexes, with females often displaying less extreme symptoms than males [13] (also reviewed in [14] [15]). As well as presenting around 4 years after males, females have a second peak of onset around the age of 50. Animal studies suggest this is due to the protective effects of oestrogen until menopause [16] and treatment with combined oestrogen and antipsychotics has been shown to accelerate symptom remission, when compared to antipsychotic treatment alone, in both males and females.

Brain morphology of affected individuals is often abnormal with increased ventricle size and reduction in anterior hippocampal volume commonly shown by MRI scanning [17]. Patients on continual medication had less reductions in hippocampal volume over time, and those with a longer duration of psychosis prior to first treatment had greater reductions in temporal lobe volume, suggesting that either the reduced volume is due to developmental effects or breakdown during the initial onset of psychosis. It is proposed that the increases in ventricle size are not due to confounding treatment effects [18] as patients with better disease outcome and drug response had less changes in ventricular size. Recent meta-analysis studies of brain morphology in schizophrenia have found reductions in grey matter volume in frontal, temporal, insular and cingulate cortex and thalamus [19]. Overall cerebral volume is reduced in schizophrenia with an increase in lateral ventrical volume and bilateral reduction in amygdala and hippocamal volume relative to the overall decrease in brain size [20]. Further meta-analysis has also shown bilateral reductions in thalamic volume in both fist-episode and chronic schizophrenia patients compared to control subjects [21]. In addition, reductions in frontal white matter

have been observed through voxel based morphometry and have also been confirmed by meta-analysis [22]. MRI studies of high risk children and non-psychotic relatives of individuals with a schizophrenia diagnosis reveal volumetric abnormalities of the prefrontal and temporal regions to a lesser extent than affected individuals, but still significantly different to controls. Further high risk studies indicate reductions in amygdala and hippocampal volumes and in superior temporal gyrus suggesting these volumetric changes may be indicative of disease susceptibility (reviewed in [23]).

The main treatment for patients with schizophrenia is anti-psychotic drugs. Classic treatments involve the use of Thorazine plus chlorpromazine and Haloperidol which are effective in treating the positive symptoms while newer 'atypical' (second generation) treatments such as Risperidone and Clozapine are also partially effective in reducing the negative symptoms [24]. Clozapine acts by blocking receptors for some key neurotransmitters in the brain, including dopamine type 4 receptors, serotonin type 2 receptors, norepinephrine receptors, acetylcholine receptors, and histamine receptors [25]. It is also a partial blocker of the dopamine type 2 receptor. Typical (first generation) antipsychotics such as Haloperidol act primarily on the dopamine type 2 receptor and can have sedative effects. Approximately one third of patients do not respond to treatment with first generation antipsychotics. Clozapine was the first second generation antipsychotic to be licenced by the FDA for treatment of treatment resistant schizophrenia [26] and is currently the only effective treatment for refractory schizophrenia. The main side effect of Clozapine use is the development of agranulocytosis, a reduction in the white blood cell count, which can be fatal due to the high risk of infection resulting from a suppressed immune system [27]. Risk can be dramatically reduced through monitoring of white blood cell counts. Side effects of treatment with first generation antipsychotics include dysphoria, dystonia, akathisia, dyskinesia, and Parkinsonian motor symptoms. There may also be an increase in the depressive/demoralization aspects of illness course, impaired learning, and slow information processing, and increased hostility, aggression, and suicide [28-30]. While the majority of side effects seen in the first generation antipsychotics are not present in

the second generation antipsychotics, they do carry their own risks. These include metabolic syndrome, hyperlipidemia and reduced insulin sensitivity with an increased incidence of diabetes. A reduced lifespan due to exposure to adverse drug effects is also observed [31-33]. Current work is concentrating on the discovery of drug targets outwith the dopamine system, that may reduce the negative and cognitive symptoms not helped by the current anti-psychotics used [34] and may have more favourable side effects than current treatments. The use of multiple antipsychotics and/or augmentation strategies (such as combined treatment with estrogen in males [35]) are also currently being studied [36]. Gender differences in presentation of schizophrenia have been observed for a long time, and it is proposed that estrogen and other female sex hormones may have a neuroprotective effect [37], delaying presentation of symptoms in females til later in life when hormonal changes occur due to menopause [38]. Psychotherapy is occassionally used along-side a programme of drug treatment in schizophrenia patients but, however effective these treatments may be in controlling symptoms, there is currently no cure for schizophrenia and treatment is lifelong.

Like schizophrenia, bipolar disorder has a lifetime prevalence of approximately 1% and affects an equal proportion of males and females [39]. Bipolar disorder is a severe mood disorder characterised by mood disturbances ranging from severe depression to extreme elation and is often accompanied by psychotic features and cognitive changes. Pathological mood swings may occur spontaneously or be cyclic in nature while some individuals will have predominant manic or depressive episodes with few mood swings. In some cases states can be mixed, with characteristics of mania intruding upon the depressive episode or vice versa. When in the depressive phase sufferers will, for example, display social withdrawal, anhedonia, difficulty with concentration and decision making, fatigue and loss of self esteem. The prevalence of suicide in bipolar individuals is high when in the depressive phase and many require constant monitoring and care to prevent self injury [40]. As an almost a direct result of this, 35% of the £200 million a year spent on bipolar disorder by the NHS is the result of hospital admissions alone [41]. In contrast, while in the manic phase sufferers will display elevated mood, delusions, hallucinations, extreme restlessness and inflated self esteem. Males and females do not

display sex differences in age of onset (mean age ~ 25) but show differences in type of episode at first presentation. Males are more likely to present with mania prior to initial diagnosis while females more commonly present with major depression [42]. Bipolar patients normally receive treatment in the form of lithium and other mood stabilisers. Lithium is the most favoured of the mood stabilisers as it does not have a sedative effect while still being effective, and is effective for both manic and depressive phases [43]. The exact mechanism by which lithium works is still unknown but studies of animal models suggests a neuroprotective role against glutamate induced excitotoxicity [44]. Psychotherapy is also widely used as treatment for bipolar disorder.

Recurrent major depression carries a lifetime risk of 16.2% [45] and often associates with schizophrenia or bipolar disorder in high risk families [5, 8]. Recurrent major depression is unipolar and characterised by phases of severe depression without mania. Symptoms are similar to those of bipolar individuals in the depressed phase. Treatment of recurrent major depression with anti-depressant drugs is somewhat effective (around 53% improvement rate) [46], however, residual symptoms including insomnia, fatigue and anhedonia persist in many patients [47], increasing the risk of full relapse [48]. Overall one third of patients will make a full recovery, one third will have persistent reduced symptoms and one third do not respond to treatment.

1.2 A Genetic Basis for Major Mental Illness

The definitive causes of schizophrenia and severe mood disorders are as yet unknown, but it has been shown that genetic control is involved [49]. Familial studies have shown that an individual's risk of developing schizophrenia increases relative to how closely related they are to a sufferer [50]. Adoption studies have shown that children with one schizophrenic parent have around a 13% chance of developing the illness later in life (compared to 1% in the general population) even when raised by mentally healthy adoptive parents [51]. It has been claimed that the heritability estimates are incompatible

with transmission through a single major locus in the majority of cases, but become plausible when modelled with multiple loci [52].

Bipolar disorder is also heritable; concordance rates in monozygotic twins are ~43% while in dizygotic twins this drops to ~6% [53]. Around half of patients with bipolar disorder have a parent with a major mood disorder [54]. Family studies have shown that families with probands of schizophrenia have a higher prevalence of unipolar depression, and aggregation of psychotic affective disorders occurs in families of probands with either schizophrenia or bipolar disorder [55].

In 2009, Lichtenstein *et al* [5] carried out a large scale study of the Swedish population to asses environmental and genetic factors involved in the predisposition to schizophrenia and bipolar disorder. Using multivariate generalised linear mixed model analysis of information from the multi-generation register and the hospital discharge register, they could investigate these risk factors for schizophrenia and bipolar disorder and the comorbidity of the disorders. They found an increased risk of schizophrenia or bipolar disorder in individuals with a first degree relative who suffered from the same disorder. The risk was higher in full siblings than half siblings with heritability estimates of 64% for schizophrenia and 59% for bipolar disorder. There was also an increased risk of schizophrenia in relatives of probands with bipolar disorder, including in children adopted and raised by healthy parents. Common additive genetic effects contributed highly to the comorbidity between the disorders (63%). Patterns of inheritance of major mental illness are not simple and it is likely many interacting factors are involved. Large families that display quasi-Mendelian segregation do exist [56], which highlight potential genetic risk factors of major effect conferring susceptibility to major mental illness in some cases. Both genetic and allelic heterogeneity occur in major mental illness. Genetic heterogeneity accounts for the large number of genes that give rise to the same classification of mental illness. Allelic heterogeneity occurs when multiple different single mutations affecting the same gene give rise to the same disorder, such as the case of *DISC1*, described in section 1.3.

1.2.1 Linkage Analysis

Linkage analysis involves the study of closely related, high risk families (ie families with multiply affected members where genetic factor(s) show high penetrance), to identify a region of the chromosome that co-segregates with the disease. This information can then be used to study genetic mutations within the linkage regions to identify potential candidate genes. Linkage analysis tends to identify large regions of the genome due to the sharing of large sections of the chromosome between subjects, hence the subsequent identification of sucseptibility genes can be difficult.

None the less, several linkage regions identified for schizophrenia and bipolar disorder have been replicated in multiple studies. 1p13-q23, 1q42, 2p12-q22, 5q21-q33, 6p24-p22, 6q16-q25, 8p22-p11, 8p22-p21, 10p15-p11, 13q32-q34 and 22q11-q12 [57, 58] are now well established linkage regions for schizophrenia, while 5p15, 6q21, 8q24, 10q26, 11p15, 12q24, 17q25, 18p11, 18q22, 20q13, 22q12 and Xq26 [59] have been established as strong linkage regions for bipolar disorder. What is most evident from linkage studies to date, is the large number of putative susceptibility regions that have been identified, scattered throughout the entire genome, which supports the theory of multiple genes of small effect [60].

1.2.2 Association Studies

Classical association analysis differs from linkage analysis as it involves the study of unrelated individuals in a population for the frequency of alleles (particularly SNPs – single nucleotide polymorphisms). Up until recently, individual functional candidate genes or in positional candidate genes within identified linkage peaks were studied. The completion of the human genome sequence has greatly facilitated the identification and study of genes for complex human diseases. In addition, the International HapMap project (www.hapmap.org), which characterises SNP frequency and correlation in samples from four geographically diverse populations, has allowed the identification of patterns of linkage disequilibrium within and between populations (The International HapMap Consortium 2003). The basis of the HapMap project is that within populations,

the human genome is divided into linkage disequilibrium regions (LD) between hotspots of recombination [61] and that the extent of LD will determine the tendancy of alleles at two or more loci to be inherited together at a higher degree than that expected by chance [62]. In numerical terms, this is the difference between observed and expected allelic frequencies. LD is influenced by rate of mutation, rate of recombination, genetic drift, population structure and genetic linkage. The use of LD to define tagging SNPs can, therefore, cover more of the variation within the region of interest than the genotyping of randomly spread SNPs [63]. Association studies for schizophrenia have identified *neuregulin 1 (NRG1)* and *disrupted in schizophrenia 1 (DISC1)* as strong candidate genes ([64-66] and others). Originally identified by linkage studies they have been supported by findings from classic association studies and copy number variants, suggesting particular polymorphisms confer risk population wide and are not confined to distinct family groups. mRNA analysis confirms the expression of these genes in distinct brain regions involved in the pathology of schizophrenia, including the prefrontal cortex and hippocampus, and functional data links many of these genes to processes which are thought to be involved in major mental illness[67].

The combination of commercial high throughput genotyping and knowledge of linkage disequilibrium means that association studies can now be carried out in a genome wide fashion, and are no longer restricted to distinct genes or chromosomal regions. GWAS (Genome Wide Association Study) combines classical association analysis with the positional cloning flexibility of genome-wide linkage scans to provide a powerful tool for the identification of candidate genes, with genome wide significance that requires no prior hypothesis about the role or function of the candidate genes in disease pathology [68]. While some possible susceptibility genes for bipolar disorder have been identified by classical association studies, the GWAS method has generally been more successful for schizophrenia. To date, the largest schizophrenia GWAS tested approximately one million SNPs in 3322 schizophrenia and 3587 controls [7]. They found association with *MYO18B* and 450 SNPs in the major histocombatibility complex in schizophrenia cases. The major histocompatability complex had also been implicated by Steffanson *et al* (2009) [69] using combined SNP data from several large scale genome wide scans. Kirov

et al (2008) [70] carried out a GWAS study using parent-offspring trios, to minimise the risk of population stratification, with 574 schizophrenia patients, their parents and 605 unaffected individuals and identified *RBP1* and *CCDC60* as susceptibility genes for schizophrenia. *RBP1* inhibits PI3K/Akt signalling which has previously been implicated in schizophrenia pathology [71]. In addition, *IL3RA*, *CSF2RA* and *RELN* have also been associated with schizophrenia [72-74]. In 2010, one GWAS study tested 572888 markers for association with schizophrenia and found no genome wide significance in the sample set from a Norwegian population [75] but found significance when expanded to a large European cohort. What is apparent from these studies is that there are multiple genes involved and little overlap of genes identified between sample sets. This apparent lack of reproducibility could be due to a number of factors, including the use of different SNP sets, genotyping platforms, sample size and genetic or phenotypic heterogeneity between samples. This second factor was addressed in part by the Wellcome Trust Case Control Consortium in 2007. They examined seven common diseases, including bipolar disorder, and demonstrated the use of a carefully selected and consistant control group was crucial in the reproducibility of results [76].

The Psychiatric GWAS Consortium (PGC) was set up in 2007 with the aim of consolidating the data from GWAS studies by statistically robust meta-analysis. To date there are five disease working groups (schizophrenia, bipolar disorder, major depressive disorder, autism and ADHD), a statistical analysis group and a cross disease working group. Early studies proved that small sample sizes (<1000 cases and 1000 controls) were often not large enough to reach genome wide significance and that larger samples (around 3000 cases and 3000 controls) were required for greater power. The formation of the PGC has allowed researchers access to far more sample sets and enabled them to increase the power of their analysis. The four main aims were to harmonize data, carry out within disorder meta-analysis, increase availablitiy of shared data and to identify convincing associations common to two or more of the five disease groups [77]. A number of genes have been associated with schizophrenia and bipolar disorder through GWAS meta-analysis. Zinc-finger protein 804A (*ZNF804A*) was originally associated with schizophrenia and bipolar disorder through genome wide association. Further meta-

analysis has replicated this association with high statistical significance ($p=2.5x10^{-11}$ schizophrenia, $p=4.1x10^{-13}$ schizophrenia and bipolar disorder combined) supporting the theory of genetic overlap between disorders [78]. Additional meta-analysis (n=5142 cases/6561 controls) have identified a SNP mapping 85kDa from fibroblast growth factor receptor 2 (*FGFR2*) with high association with schizophrenia (p=0.0009) [79]. The accumulation of data from multiple studies has led to the emergence of some interesting and consistent patterns. It is proposed that certain risk haplotypes in these genes may have a detrimental effect on brain function which is modulated by poorly understood environmental variables and social factors [80], and the spectrum of clinical features are suggestive of risk through overlapping sets of genes [81].

1.2.3 Copy Number Variation

Much recent work has investigated the role of copy number variants (CNVs) in neurodevelopmental disorders (reviewed in Kirov 2010) [82]. As the aetiology of schizophrenia and other major mental illness is so varied, with multiple genetic components identified, the general consensus is that multiple small components are likely to be responsible, and as CNVs account for a substantial proportion of human genetic variation, they are likely to play an important role.

A study of 418 individuals, including 150 individuals affected with schizophrenia or schizoaffective disorder and 268 healthy controls, identified 24 previously unreported copy number variants in schizophrenia cases [83]. These included deletions in *NRXN1*, *ERBB4, GRM7* and *SLC1A*, all known synaptic genes. Additionally, Xu *et al* (2008) [84] identified *de novo* and inherited copy number variations in schizophrenia cases associated with neural development, small GTPase activity and RNA binding/processing. In a study of 724 patients with psychiatric disorders and 314 healthy controls, Saus *et al* (2010) [85] studied the dosage effects of 68 known candidate genes for psychiatric conditions overlapping with CNVs. Contrary to revious reports, no statistically significant difference was found in the overall burden of gains or losses in psychiatric patients compared to control individuals. However, 47% of rare structural variants identified were found only

in psychiatric disorder patients and not in control individuals. These included variants in *GRM7*, and *COMT*, both previously implicated in schizophrenia. Need *et al* (2009) were also unable to confirm the 'load burden' previously reported but did identify a number of large deletions in cases but not controls. Using samples from 1013 cases and 1084 controls from schizophrenia cohorts in Aberdeen, Munich and America, Need *et al* found that deletions >2Mb were not present in control subjects but were in many schizophrenia cases. They propose that large deletions confer risk to psychiatric illness. In the Aberdeen cohort, deletions affecting *NDE1, MPV17L, ABCC1, KIAA0430, KIAA0866, MRP6* and *SPRY2* were found to be significantly associated with schizophrenia, and in the Munich cohort deletions affecting *TUSC3, PCM1, NAT1, NAT2* and *ASAH1* were significantly associated [86]. Many of these genes have been implicated in schizophrenia previously.

1.2.4 Genome wide gene expression analysis

From current research it is clear that schizophrenia and related synaptopathies are complex trait disorders, attributable to multiple genes and epigenetic factors [87]. Microarray technology allows relatively rapid, large scale screening of genes with the potential to identify candidate genes and pathways involved in complex trait disorders.

Recent microarray studies of human schizophrenia patients have identified multiple presynaptic and myelin related genes dysregulated in the prefrontal cortex [88, 89]. Mirnics *et al* (2000) identified a group of genes involved in presynaptic secretory function whose RNA expression was decreased in schizophrenia patients. Genes involved in GABA and glutamate neurotransmission were also reduced. These findings were consistent with previous targeted gene expression studies in hippocampus, which found GABA and glutamate receptor function was decreased in schizophrenia patients [90, 91]. Vawter *et al* (2002) [89] carried out a targeted microarray for 1127 brain related genes in the dorsolateral prefrontal cortex of post mortem schizophrenia brains and matched controls and reported an overlap in 5 of the presynaptic secretory genes, GABA and glutamate receptor genes identified by Mirnics *et al*. In a similar study, Sugai *et al* (2004) identified oligodendrocyte and astrocyte related genes, and growth/neurotrophic factors

and their receptors, as being altered in the prefrontal cortex of schizophrenia patients [92]. In the superior temporal gyrus, which has connections to the thalamus, the limbic system and the prefrontal cortex, altered expression of genes involved in neurotransmission, neurodevelopment and presynaptic function were identified in post mortem brain tissue of schizophrenia patients [93]. Two studies have also identified genes involved in mitochondrial function and oxidative stress as being differentially expressed in schizophrenia patients [94, 95].

While it is clear that microarray studies are somewhat useful in identifying candidate genes for the complex psychiatric disorders, there are a number of drawbacks to the method. Bunney *et al* (2003) [96] highlight the possible pitfalls of using microarray technology for studying psychiatric disorders. One major limiting factor is the availability of post mortem brain tissue from psychiatric patients. Post mortem interval and heterogeneity of human samples will effect gene expression and are difficult to control for. Brain tissue from subjects who had previously been medicated will display altered expression in some genes relative to unmedicated, otherwise matched, individuals. It is therefore important to note the medical history of the subject prior to making conclusions about gene expression patterns [96]. An alternative approach using samples from live patients, such as lymphoblastoid cell lines derived from blood samples, can increase the availability of subjects and allow ongoing assessment. Lymphoblastoid cell lines overcome the problems associated with drug treatments and post mortem interval, however, many genes that have been identified by studies using brain tissue are not expressed in blood [97] and studies comparing gene expression in brain and that in blood suggest the different tissue types will yield very different results. It is therefore suggested that while brain tissue can provide candidate genes, blood samples will provide possible biomarkers for disease.

It is of most interest that the same genes appear to recur in genetic studies of schizophrenia and mood disorders suggesting a causative link between the two illnesses [98]. The genetic overlaps between schizophrenia, recurrent major depression and bipolar disorder are becoming increasingly more convincing through the results of high

throughput microarray studies [99] and some researchers have suggested they may not be separate disorders but variations on a spectrum of a single disorder[98].

1.2.5 Cytogenetics

Cytogenetics is the study of chromosomal abnormalities that cause disease. This can be the addition of an entire chromosome (as in Down's syndrome), the absence of a chromosome (Turner's syndrome) or the translocation of one part of a chromosome to another. Diseases resulting from chromosomal abnormalities tend to be Mendelian in nature, however a number of genes for major mental illness have been discovered through the use of cytogenetics [100] suggesting that single genes of large effect do exist for these disorders, and that the polygenic theory of psychiatric illness may not be the whole story. This project focuses on the *DISC1* gene, which was first identified in a balanced translocation in a large Scottish family with multiple individuals affected by schizophrenia.

1.3 Identification of Disrupted in Schizophrenia-1

A Scottish family was first reported by Jacobs (1970) [101] who noted a balanced translocation – (t1:11) (q42;q14.3) – in an individual with an adolescent conduct disorder and subsequently in members of four generations of the extended family. A 20 year follow up study [102] observed increased occurrence of major psychiatric disorders including schizophrenia, bipolar disorder and major depression in translocation carriers, but not in non-carrier relatives (LOD=7.1)[8]. Of interest was one individual diagnosed with bipolar disorder whose child went on to develop schizophrenia, adding to the evidence that the functional psychoses have overlapping genetic predispositions. Further studies determined that the translocation breakpoint directly disrupts two genes on chromosome 1 (*DISC1* and *DISC2*) [56] making them ideal candidates for further research. This translocation breakpoint occurs in intron 8 of *DISC1* leading to the transfer of the coding sequence for 257 C-terminal amino acids (exons 9-13) from chromosome 1

to chromosome 11 in translocation carriers[103]. There is, however, no evidence of a C-terminal truncated protein in translocation carriers [104]. Instead the expression of DISC1 in the brains of affected individuals is reduced by half compared to karyotypically normal controls, suggesting haploinsufficiency is the most likely mechanism of susceptibility to schizophrenia and other major mental illness in (t1:11) (q42;q14.3) translocation carriers. Zhou *et al* [105] later identified a novel gene that was disrupted by the translocation on chromosome 11. This gene (termed Boymaw) was suggested to form fusion proteins with DISC1 in translocation carriers. While these fusion proteins have been generated and their function studied in cell cultures [105], work to determine whether fusion proteins are transcribed in translocation carriers is still ongoing.

Translocation carriers (as well as unrelated patients with schizophrenia without the translocation) show prolonged latency and reduced amplitude of P300 event related potential (a measure of attention dependant information processing which has been shown to be impaired in schizophrenia patients) compared with controls, and karyotypically normal relatives[8, 103, 106]. This finding led to the proposal that the balanced (t1;11)(q42.1;q14.3) translocation conferred predisposition to slower and more inefficient processing of stimuli in short term memory. Thus *DISC1* is likely to participate in pathways important for cognitive function. *DISC2* is antisense to *DISC1,* and it is also disrupted by the translocation. Sequence analysis of *DISC2* suggests it is transcribed in the opposite direction to *DISC1* and has no protein coding potential. It appears to be an anti-sense RNA gene, which may regulate activity of *DISC1* [103]. So far *DISC2* has only been detected in humans, suggesting relatively recent evolutionary origins.

While the translocation is confined to one family, many association and linkage studies of other populations have discovered other variants within the *DISC1* locus that also associate with schizophrenia, bipolar disorder and recurrent major depression [8, 107-110]. Three independent studies have, however, failed to find association between *DISC1* locus variations and schizophrenia [111-113] although it is proposed that this can be

explained by small sample size, too few SNPs being studied and ethnic differences in susceptibility [113].

Replication of linkage in a Finish sample has shown under-transmission of three common *DISC1* haplotypes and major mental illness (HEP2, HEP3 and HEP4), and the over-transmission of one haplotype (HEP1). The under-transmitted haplotypes represent regions spanning intron 1 to exon 2, exon 4 and exon 13 respectively. HEP3 occurred in the control group at a frequency of 1% and a frequency of 8.8% in the schizophrenia patients group [114]. HEP3 is also associated with poor visual working memory and attention [115] and HEP1 has been associated with impaired long-term verbal memory and total hippocampal volume reduction [116]. In addition, the HEP3 haplotype is strongly associated with schizophrenia in a Scottish case-control group [117] and also shows overlap with a three-SNP haplotype associated with Aspergers syndrome in cohort of families with infantile autism and Aspergers syndrome [118]. This study also found association with an intragenic single nucleotide polymorphism (SNP) of *DISC1* (rs1322784) and Aspergers syndrome. They also established association between autism and a *DISC1* intragenic microsatellite (D1S2709) suggesting a strong role for *DISC1* in early onset neuropsychiatric conditions and neurodevelopment. In a genome wide linkage scan of schizoaffective disorder, Hamshere *et al* (2005) [119] reported the highest linkage peak (LOD 3.54) occurring at 1q42, close to the *DISC1* locus. Furthermore, a missense allele in exon 9 of *DISC1* was reported to be overrepresented in patients with schizoaffective disorder [120] suggesting *DISC1* may confer risk to this disorder.

DISC1 has been associated with neurocognitive functioning [121] in measures of verbal working memory and rapid visual searching. The association of variations in *DISC1* with cognitive function was investigated in a large birth cohort tested using a general mental ability test at ages 11 and 79 [122]. Cognitive ability was compared with genotype within age groups, and between ages 11 and 79 to test for effects on cognitive ageing. After adjustment for cognitive ability at age 11, it was found that females homozygous for the *DISC1* Cys (Ser704Cys SNP) allele had significantly lower cognitive ability score than males, suggesting a sex specific role of *DISC1* genotypes on cognitive ageing [122].

Callicott *et al* (2005) also report that a nonsynonymous SNP in exon 11 (Ser704Cys) of *DISC1* is associated with schizophrenia. Using fMRI they observed abnormal engagement of the hippocampus during cognitive tasks, and a reduction in hippocampal grey matter volume [123]. Further to this, Di Giorgio *et al* (2008) [124] investigated the effects of the polymorphism on grey matter volume and formation, hippocampal formation and functional coupling during memory recognition tasks. Using fMRI they showed that individuals with two Ser alleles had greater engagement of the hippocampus and greater functional coupling to the frontal cortex during memory recognition tasks, but in contrast with Callicott *et al*, observed higher grey matter volume than heterozygous individuals. While these results are in part contradictory, they both support a role for *DISC1* in hippocampal formation and function. The authors suggest this inconsistency is due to the 'flip-flop' effect, where the direction of the allele association may be dependent on interaction with another allele at a different locus, thus the same allele could exert different effects in different populations depending on their genetic pressures. The Cys allele has also been shown to reduce activation of the left medial and superior frontal gyrus during a verbal fluency task, relative to Ser homozygotes [125]. No brain regions were found to be significantly more activated in Cys carriers. These results support a strong physiological effect of the *DISC1* polymorphism in both the hippocampus and the frontal cortex. At the molecular level, sRNAi knockdown of *DISC1* results in suppressed phosphorylation of ERK and Akt. The Ser704 allele phosphorylated ERK to a greater extent than the Cys704 allele, implicating reduced biological activation of ERK due to polymorphisms in *DISC1* in major mental illness pathology [126].

Current estimates suggest that approximately 2% of individuals with schizophrenia carry a *DISC1* missense mutation [127]. As discussed in the next section, DISC1 acts as a hub protein, with multiple interactors, and as such this estimate may be under representative. Interplay between SNP variants [128], common cis-variants [128], and interacting proteins [129] has been shown to result in modest reductions of *DISC1* gene expression, which may exert subtle effects on neurodevelopment, neurophysiology and neural circuitry. The DISC1 protein has also been shown to interact with transcriptional modulators of cAMP signaling, cytoskeletal, synaptogenic, neurodevelopmental and

sensory perception proteins [128]. Current targets of psychiatric drug development are highly enriched in this group [128]. With recent evidence suggesting there may be as many as 50 *DISC1* mRNA isoforms [130], showing dramatic differences in their brain expression profiles, it is proposed the combination of common variants of low penetrance, and rare variants of high penetrance within the *DISC1* pathway may contribute a much larger fraction of the genetic variance in schizophrenia and related disorders than previously thought [127].

1.4 The Biology of DISC1 and its interactions

DISC1 acts as a molecular scaffold and has been shown to interact with multiple proteins including NUDEL and PDE4B, suggesting roles in neurodevelopmental and signalling pathways [131, 132]. The known isoforms of *DISC1* are produced by alternate splicing of 13 major exons, of which exon 2 is the longest (955bp) and present in all isoforms. Exon 2 encodes most of the protein head domain, and haplotypes spanning this region have been shown to associate with schizophrenia and working memory [107, 108].

Recently an interaction network of 127 proteins and 158 interactions has suggested DISC1 may also be involved in synapse function and development through these complex interactions [132]. The majority of these interactors can be loosely classified as cell cycle, signal transduction, cytoskeleton, intracellular transport and central nervous system development genes, clearly demonstrating the importance of DISC1 in multiple diverse and critical brain functions. A direct role for DISC1 in early brain development *in vivo* was first demonstrated by Kamiya *et al* (2005) [133] who reported that *in utero* application of short hairpin loop RNA oligonucleotides (shRNA) targeting mouse *Disc1* can repress expression of the gene, resulting in reduced migration of neurons out of the sub-ventricular zone to the cortical plate. This is accompanied by altered cell polarity and reduced dendritic arborisation. This correlates well with evidence from mouse studies where *Disc1* is upregulated at crucial developmental time points both embryonically (E13), when neurogenesis and cell migration commence, and at the onset of puberty

(P35) when synaptogenesis is again occurring [134]. From puberty onwards the expression of *Disc1* in the mouse remains steady. Interestingly, in adult hippocampal neurogenesis, downregulation of *Disc1* results in increased acceleration of neuronal integration, resulting in mispositioning of dentate granule cells [135]. Newborn *Disc1* knockdown dentate granule cells display increased excitability, dendritic development and synapse formation. Frequency of GABAergic and glutamatergic spontaneous synaptic currents in the *Disc1* knockdown cells was consistently higher, indicative of increased synaptic excitability[135]. This would suggest that *Disc1* has distinct functions during development that differ pre and post-nataly, and that it is likely to play a role in sustained and simultaneous firing of neurons.

While DISC1 has been shown to have multiple interactors, only around 25% of these have been confirmed by further study. Of those which have been confirmed however, many have been implicated in schizophrenia and related disorders in their own right, including NDE1, NDEL1, PDE4B [65] PCM1 [129] and FEZ1 [136], and it has been suggested that in some cases where DISC1 is not directly implicated, variants acting on its binding partners may act instead to confer risk of major mental illness [65]. In the following paragraphs I will discuss the major confirmed interactors of DISC1 with links to psychiatric illness; NDE1, NDEL1, LIS1, GRB2, PCM1, GSK3β, FEZ1, ATF4 and ATF5.

NDE1, NDEL1 and LIS1 form part of a conserved nuclear distribution pathway involved in neurogenesis and neuronal migration [137-140]. Mutations in *Lis1* lead to deficits in neuronal migration, neuroblast proliferation and cortical layering [141, 142]. Similarly, mice with mutations in *Nde1* and *Ndel1* show defective neurogenesis and neuronal migration [139, 140]. DISC1 and NDEL1 have been shown to be selectively upregulated during neurite outgrowth in PC12 cells differentiated with neuronal growth factor (NGF) (Kamiya 2006) [143]. Inhibition or disturbance of the DISC1-NDEL interaction has been shown to inhibit neurite outgrowth in cell culture. This would suggest the resulting complex is crucial for successful neurite outgrowth [143].

These genes have also been associated with schizophrenia in their own right. mRNA studies of post mortem brain implicate both *LIS1* and *NDEL1* [144] in schizophrenia, while *NDE1* and *NDEL1* have also been associated with schizophrenia in a Finnish cohort [145, 146].

GRB2 competes with NDEL1 for binding to DISC1 [147] giving two functionally distinct interactions. The DISC1/GRB2 interaction is required for neurotrophin induced axonal elongation [147]. GRB2 has been shown to bind several brain proteins including PDE4D [148], huntingtin [149], amyloid precursor protein [150] and presenilin 1 [151] as well as dopamine receptors 3 and 4 [152]. As such, its involvement in Alzheimers disease and molecular pathways involved in other psychiatric syndromes has been suggested [153]

PCM1 is a centrosomal protein required for the targeting of multiple proteins to the centrosome for the regulation of microtubular dynamics [154]. Co-immunoprecipitation in HEK-293 cells suggests PCM1 and DISC1 may interact to regulate neuronal migration. In addition, knockdown of *DISC1* expression reduced accumulation of PCM1 at the centrosome. Further RNAi studies of both *DISC1* and *PCM1* show delayed radial neuronal migration if either protein is knocked down [133, 155]. Gurling *et al* (2006) [156] reported an association between haplotypes of *PCM1* and orbitofrontal grey matter defects in schizophrenia family and case-control samples although this has as yet failed to replicate in other samples.

GSK3ß is a presynaptic protein kinase that is proposed to be involved in the negative regulation of synaptic vesicle fusion events [157]. GS3Kß has been shown to suppress long-term potentiation and presynaptic release of excitatory glutamate in neurons. Mao *et al* (2009) [158] observed that suppression of *Disc1* by shRNA in embryonic mouse brains resulted in reduced neural progenitor proliferation, that could be rescued using the GSK3ß specific inhibitor SB216763. GSK3ß is inhibited by direct DISC1 interaction suggesting that a loss of DISC1 results in increased GSK3β and leads to reduced neural progenitor proliferation, leading to premature cell cycle exit and differentiation.

Therefore it is suggested that the DISC1-GSK3ß interaction is required for successful neural proliferation.

Many current typical and atypical antipsychotics increase phosphorylation of GSK3ß and enhance AKT signalling though GSK3, or by activating AKT directly. AKT1 itself has been tentatively associated with schizophrenia [159] although further studies are required. The actin binding protein Girdin (KIA1212) was initially identified as a potential DISC1 interactor by Camargo et al 2007 [132]. Since then, its interaction with DISC1 has been confirmed both *in vivo* and *in vitro*. Interaction of DISC1 with Girdin in HEK293 cells suppresses AKT signaling [160]. When Kim *et al* [160] knocked down *Disc1* in mouse hippocampal neurons, Akt activation increased, and increased dendritic growth and abnormal migration were observed. Similar observations were made when *Girdin* was selectively overexpressed in these neurons [160]. Interestingly, the effects of *Disc1* suppression could be rescued by rapamycin, which inhibits an AKT activated effector pathway. Enomoto et al (2009) [161] showed that *Girdin* suppression also mimicked the effects of *Disc1* supression in cultured hippocampal neurons. Girdin appears to be anchored at the growth cone by DISC1, and regulates the migration and positioning of dentate gyrus cells from this position. Girdin has also been shown to interact with NDEL, another known interactor if DISC1.

Fasciculation and Elongation factor Zeta 1 (FEZ1) is involved in axon outgrowth and has been shown to be expressed in a pattern similar to that of DISC1 in the rat brain [162]. In addition, protein overexpression leads to robust binding of DISC1 to FEZ1 and co-localisation to growth cones in rat cultured hippocampal neurons [163]. Expression of FEZ1 in post-mortem brain samples is reduced in subjects with schizophrenia [144] however genetic/association evidence for FEZ1 as a susceptibility gene for schizophrenia is weak [164-166].

ATF4 and ATF5 have been associated with schizophrenia through the use of yeast two hybrid screens of DISC1 [167]. ATF4 has been shown to interact directly with nuclear DISC1 and a corepressor, N-Cor, to modulate CRE-mediated gene transcription [168].

There is currently little evidence to suggest ATF4 and ATF5 are involved in psychiatric illness in their own right. As they are both cAMP-resposive transcription factors, however, and because ATF4 is partially regulated by *NRG1*, another well characterised schizophrenia susceptibility gene [169] they are of interest for further study. Furthermore, it has been shown that DISC1 and PDE4B interact dynamically to regulate cAMP signalling [131]. Disruption of this interaction, as predicted by some mouse models (section 1.7) would, therefore, be predicted to alter modulation of cAMP signalling and may result in abnormal gene transcription. PDE4s are orthologous to the Drosophila *Dunce*, which is involved in learning and memory and known to affect synaptic plasticity, which in turn requires alteration of gene expression profiles [170, 171]. DISC1 also possibly binds to chromatin remodelling factors, such as SMARCE1 [132]. These interactions with transcription factors, together with the fact that DISC1 localises to the nucleus [172, 173] is consistent with a role for DISC1 in transcriptional regulation. These are discussed further in the next section.

1.5 Current hypotheses of Major Mental Illness

1.5.1 The Neurotransmitter Hypotheses

Selective targeting of individual neurotransmitters has not yet yielded many therapeutic results bringing us to the obvious conclusion that no single neurotransmitter hypothesis can account for the wide range of symptoms present in schizophrenia. However, the combined effect of neurotransmitter dysfunction is likely to account for a large proportion of disease pathology. The serotoninergic system has been implicated in both schizophrenia and major depression. While no abnormalities in receptor density have been found in schizophrenia cases [174], the affinity of successful atypical antipsychotics for the serotonin receptor 5-HT2A support a role for the receptor in schizophrenia pathology [175]. The formation of the glutamate hypothesis of schizophrenia came from the observation that PCP, an NMDA receptor antagonist, could induce schizophrenia-like symptoms in healthy subjects [176]. This is also true of other NMDA antagonists [177-179]. An increase in NMDA receptor binding has been observed in the cortex of post-

mortem brain samples from schizophrenia patients [180] and a decreased release of glutamate has been reported in synaptosomes prepared from frozen brain samples of schizophrenia patients [181]. Genes involved in NMDA receptor function have been associated with schizophrenia adding further weight to the glutamate hypothesis. One of the most established associations, NRG1, regulates the expression of NMDA receptors through ErbB4. GRIN2B, the gene coding for the NR2B subunit of the NMDA receptor also shows a small but significant association with schizophrenia [182].

Dysfunction of dopamine neurotransmission has also been implicated in the pathology of schizophrenia based on the fact that dopamine-mimetic drugs elicit hallucinations, and that neuroleptics caused rigidity [183]. The original hypothesis stated that dopamine pathways may be overactive in schizophrenia. Recent studies have suggested an increase in dopamine 2 receptors in the striata of schizophrenia patients, which, while not statistically significant, indicated a substantial positive trend in dopamine activity [184] [185]. In addition, a transgenic mouse with selectively inducible over-expression of dopamine 2 receptors displays selective cognitive impairement in working memory tasks, but does not exhibit a global cognitive deficit [186]. This working memory deficit is not reversed when the transgene is switched off indicating that it is the expression level during development and not continued expression that is of key importance. Deficits in spacial working memory in individuals with schizophrenia have been well characterized ([187, 188] and others)

1.5.2 The Developmental Hypothesis

The neurodevelopmental hypothesis states that small neurodevelopmental deficits in key circuitry during brain development could lie dormant until puberty, when normal molecular changes could facilitate the onset of disease in affected individuals [189]. Studies of children from high risk family groups suggest that delays in motor and neurological development, deficits in attention and verbal short-term memory and poor social skills are evident in children who later go on to develop schizophrenia [190]. Individuals with high genetic liability have been shown to have reductions in grey matter

density in the prefrontal and temporal cortex, amygdala, hippocampus and superior temporal gyrus (reviewed in [191]). High risk individuals with subsequent follow up scans revealed a significantly greater reduction in temporal lobe size in those who developed psychotic symptoms, suggestive that the reduction in grey matter volume may be an early indicator of disease. It has also been reported that early onset human schizophrenia patients display a delay in brain development with respect to overall brain volume [192]. Evidence from animal models of schizophrenia identifies clear neurodevelopmental deficits in migration and proliferation of neurons in the hippocampus and cortex [133, 193], two regions implicated in schizophrenia pathology. Some very compelling evidence from transient knock-down studies of *Disc1* suggests that a loss of *Disc1*, particularly in pyrimidal cells of the prefrontal cortex, results in behavioural abnormalities after puberty [194]. There was a reduction in parvalbumin-positive cells (a marker of fast-spiking GABA interneurons) and reduced dopamine expression after puberty but not before. Knock down mice also displayed a marked deficit in prepulse inhibition, which was not observed at earlier ages. The authors suggest that a lack of Disc1 during development results in abnormalities in the postnatal maturation of dopaminergic neurons which causes dendritic abnormalities and an overall disturbance in neural circuitry evident after puberty [194].

1.5.3 The cAMP hypothesis

PDE4B and *PDE4D* belong to the family of phosphodiesterase encoding genes whose protein products are involved in the inactivation of adenosine 3',5'-monophosphate (cAMP) in the cell. cAMP is part of a second messenger system thought to be critically involved in learning and memory, and mood. Induction of cAMP carried out by Millar *et al*[104] resulted in decreased binding of PDE4B to DISC1 suggesting that this interaction is cAMP dependent. Immunoprecipitation of PDE4B showed that binding to DISC1 was direct and that it required an intact DISC1 N-terminal "head domain". Therefore it is possible that functional or structural variation in DISC1 will modulate the interaction with PDE4B, affecting cAMP catabolism and thus altering expression of other downstream genes. In addition, a chromosomal rearrangement disrupting PDE4B was

identified in a family with schizophrenia and related psychoses [131]. PDE4 activity has been linked to memory formation in the fly (cognate *dunce* mutation) and to mood disorders in the mouse [195]. It is also the target for rolipram, a potent prototypic antipsychotic and antidepressant [196]. Furthermore, Clapcote et al [197] describe two mouse models of major mental illness that both have mutations in a Pde4b binding site of Disc1.

1.6 Genome-wide expression and pathway analysis

Analysis of gene expression datasets using pathway and network analysis is now providing insight into the regulatory mechanisms of disease. Porteous and Hennah (2009) [128] employed pathway analysis to determine correlations between variants in *DISC1* pathway genes and levels of gene expression in public domain datasets. Data mining of the four available HapMap population cohorts (GSE6536 in NCBI GEO) [198] identified six cis-acting variants that showed association in at least three of the four cohorts. In all cases the minor allele was associated with reduced *DISC1* expression. Further analysis incorporating the six cis-acting variants, three known missense mutations, and three known genetic interplay SNPs (conferring risk, neutral or protective effects) identified 100 genes that could be connected in a pathway of interacting molecules. Gene ontology analysis of this pathway identified over-enrichment of genes involved in cytoskeletal function, transcription factor function, synaptogenesis and sensory perception. While the data from Porteous and Hennah's study pertains to normal genetic variance, its relevance to unraveling the function of *DISC1* is undeniable, and the identification of regulatory networks gives insight into the mechanisms of psychiatric disease. The use of pathway analysis to further test the validity of candidate genes and their likely contribution to psychiatric disorders is without doubt a valuable tool for gene expression analysis.

Recently, Torkamani *et al* (2010) have used co expression network analysis to determine gene clusters of functional significance in schizophrenia [199]. Genes were organised into functional modules, which were co-regulated and as such likely to be involved in

similar cellular processes and pathways. Using standard analysis of differential gene expression they found that genes with altered expression in schizophrenia clustered into five distinct gene modules, four of which were preferentially associated with neuronal function. They also found that changes in gene expression between patients with schizophrenia and controls differ with age. Harris *et al* [200] previously reported genes related to CNS development, neuron guidance and neurotransmitter secretion were down-regulated in healthy individuals during post natal development in a sample set from individuals from birth to early twenties. Torkamani *et al* propose that the downregulation observed in their control subjects is a continuation of this process in normal individuals. These changes were not observed in subjects with schizophrenia however, suggesting a progressive neurodevelopmental deficit. It is suggested that the lack of normal down-regulation of these CNS developmental genes may act as a trigger for the onset of disease [199]. As schizophrenia is not normally diagnosed until the late teens, gene expression measurements of individuals with schizophrenia during their early post-natal development is not possible. The use of animal models may be useful in shedding light on normal age-related changes in neuronal gene expression levels.

The use of animal models is becoming increasingly popular in psychiatric research and allows further investigation of candidate genes in a system less limited by sample availability and tissue type. The effect of alterations in specific candidate genes can also be assessed taking investigation into the mechanisms of psychiatric disorders forward more rapidly.

1.7 The Schizophrenia Mouse

One question I have been repeated asked is ‘so how do you know if a mouse is schizophrenic?’. This is of course a loaded question, mice are not schizophrenic. We can however genetically or pharmacologically manipulate mice to display certain behavioural phenotypes which model the human condition in an effort to further understand the underlying mechanisms, but major psychiatric syndromes are uniquely human conditions

and without a defining pathology, and as such, modeling is limited to examination of a specific dimension of the disorder [201]. Endophenotypes are heritable, intermediate phenotypes between genes and expression of a disorder and are often used as 'biomarkers' for disease. It is generally considered that each complex disorder phenotype is made up of a number of endophenotypes, which are under different genetic control, so by studying single endophenotypes the problem of finding genes involved becomes less complex [202]. For example, genes influencing liability to mental disorders such as schizophrenia are likely to act on the multiple neural systems that have already been associated with this disorder. These could include serotonergic, glutamatergic and dopaminergic systems mediating processes such as learning and memory, social cognition, emotion and sensorymotor gating. Neurophysiological, neuropsychological and biochemical measurements from these systems give quantifiable measures separate from disease symptoms that can be used to map multiple genes of small effect [202]. The use of endophenotypes also allows researchers to model certain aspects of psychiatric disorders in animals, using quantifiable measures that can be related to the human condition [203]. While there are no universally accepted risk loci for major psychiatric syndromes [204, 205], researchers are concentrating efforts on modeling loci which have been identified through cytogenetics, copy number variants, association, linkage or gene expression studies as candidate susceptibility genes. Multiple mouse models of psychiatric syndromes now exist, though this section will concentrate on examples of *Disc1* animal models.

Koike *et al* (2006) [206] reported that a 25bp deletion variant in exon six of *Disc1* in the 129S6/SvEv mouse strain, results in a lack of full length Disc1 protein and impairment of working memory in a C57BL6 backcross (>98% C57BL6 genotype). Gross brain morphology of these animals was reported to be normal, however in a delayed non-match space test (which measures special working memory) there was a clear working memory impairment in both heterozygous and homozygous mutant animals, with a dosage dependant effect. This same model was used by Kvajo *et al* [207] when studying neuronal architecture and cognition. They confirmed the previous working memory impairment and found alterations in the organisation of neurons in the dentate gyrus, and

reduction of short term potentiation of the CA3/CA1 hippocampal synapses. The strength of synaptic transmission, release probability and long term potentiation were not significantly different in the mutant animal. Working memory deficits are considered fundamental cognitive symptoms of schizophrenia [208] and impaired dendritic growth and misorientation of dendrites have previously been described in human schizophrenia subjects [209]. A third study of this mouse strain, which used all available DISC1 antibodies, found that all isoforms of DISC1 were still expressed, suggesting bypassing of the deletion through exon skipping [210]. It is yet to be confirmed whether the original working memory deficits reported by Koike *et al* are indeed a result of the *Disc1* deletion or what other factors may be in play.

Another mouse model, carrying a dominant-negative *Disc1* mutation on a *CamKII* promoter, shows enlargement of the lateral ventricles and a selective reduction in the immunoreactivity of parvalbumin in the cortex, indicating a deficit in interneurons, leading to cortical asynchrony [193]. Enlargement of the lateral ventricles has long been associated with schizophrenia [211] and is usually present at disease diagnosis. In addition, this model also displays disturbed sensorymotor gating, increased hyperactivity and measures of anhedonia, all behavioural attributes that have previously been reported in patients with schizophrenia. It should be noted that these behavioural anomalies are not restricted to schizophrenia, and have been associated with multiple psychiatric syndromes, but the combination of phenotypes and the known genetic connection between *DISC1* and schizophrenia indicate that this model may mimic some aspects of the human schizophrenia phenotype.

In 2008 Pletnikov *et al* [212] generated a transgenic mouse model with inducible expression of mutant human *DISC1*. Expression of the transgene was limited to the forebrain regions to counteract any effect of the mutation elsewhere. No neurodevelopmental abnormalities were observed, but there was mild enlargement of the lateral ventricles and primary cortical neurons derived from mutant *DISC1* mice produced significantly less elaborate neurite outgrowth compared to those from wild-type controls. Some gender specific behavioural phenotypes were observed which were consistent with

other *Disc1* mutant mouse models. Inducible *Disc1* male mice show increased horizontal activity in the open field similar to the dominant-negative *Disc1* model [193] and the L100P ENU mouse model [197]. Males also display reduced social interaction consistent with the Q31L ENU mouse model [197], and increased aggression. Females display a reduction in spacial learning and memory, which has not previously been shown. Further studies indicate that prenatal over-expression of N-terminal human DISC1 results in a reduction of parvalbumin-positive and dopamine-positive neurons in adult mouse cortex. This was indistinguishable to that observed in animals with continuous expression of N-terminal human DISC1, indicating that there may be a critical period for *DISC1* expression [213].

In an effort to closely mimic the genetics of the Scottish (t1:11) translocation family, Shen *et al* (2008) developed a transgenic *Disc1* mouse that expressed a truncated form of *Disc1* comprising of exons 1-8 [214]. Truncated *Disc1* is expressed in the cerebellum, cerebral cortex and hippocampus of these mice. Enlargement of lateral ventricles and a reduction in cerebral cortex was observed along with reduced neurite outgrowth in culture consistent with observations from Pletnikov *et al* in the inducible expression model [212]. Similar to the phenotype of the dominant-negative mouse model [193] and recent work on the inducible expression mouse model [213], these animals also display a reduction in parvalbumin neurons in the cortex. This model also displays many novel phenotypes not previously reported in a *Disc1* mouse model, including reduced parvalbumin-positive neurons in the hippocampus, thinning of layer II/III in the cortex and selective decrease of neural proliferation in the cortex. Some aspects of the behavioural phenotype are consistent with those observed in both the dominant-negative mouse model [193] and the ENU mouse model (Clapcote *et al* 2007, discussed next), with increased immobility in depression related tests and reduced latent inhibition (table 1.1).

	Koike et al 2006	Hikida et al 2007	Pletnikov et al 2008	Shen et al 2008	Clapcote et al 2007	
Phenotype	**129 mice**	**DN DISC1**	**Inducible DISC1**	**Truncated Disc1**	**Q31L**	**L100P**
Brain volume	NS	NS	NS	↓ MALE	↓6%	↓13%
Lateral ventricles	-	↑	↑	↑	-	-
Cerebral cortex	NS	NS	NS	↓	↓	↓
Neural proliferation	-	-	-	↓	-	-
Neurite outgrowth	-	-	↓	↓	-	-
Corpus callosum	-	-	-	↓	-	-
Parvalbumin in MPF cortex	-	↓	↓ (Ayhan et al 2010)	↓	-	-
Parvalbumin in hippocampus	-	-	-	↓	-	-
Open field: horizontal activity	NS	↑	↑ MALE	NS	NS	↑↑
Open field: vertical activity	NS	↑	-	-	NS	↑
Anxiety	NS	NS	NS	-	NS	NS
Aggression	-	-	↑ MALE	-	-	-
Sociability	-	NS	↓ MALE	-	↓	NS
Spatial learning and memory	NS	NS	↓ FEMALE	-	NS	NS
Working memory	↓	-	-	-	↓	↓↓
Prepulse inhibition (PPI)	NS	↓	NS	-	↓	↓↓
Latent inhibition (LI)	-	-	-	↓	↓	↓↓
Immobility in forced swim test	-	↑	-	↑	↑	NS
Immobility in tail suspension test	-	-	-	↑	-	-
Stress calls	-	-	-	↓ MALE	-	-
Rolipram in PPI	-	-	-	-	NS	+++
Bupropion in PPI	-	-	-	-	+++	NS
Clozapine in LI	-	-	-	-	NS	+
Clozapine on horizontal activity	-	-	-	-	+++	++
Bupropion in FST	-	-	-	-	++	-

Table 1.1: Phenotypes of *DISC1/Disc1* mouse models (adapted from Shen et al 2008). Column 1 shows the phenotype and columns 2-7 the mouse models. NS = not significant, - = unknown, ↑ = increased, ↓ = reduced, +, ++, +++ = positive effect with $p<0.05$, 0.01 and 0.001 respectively.

Our collaborators Clapcote and Roder have developed two independent *Disc1* mouse mutants at the University of Toronto each displaying a different phenotype of psychiatric illness [197]. N-ethyl-N-nitrosourea (ENU) is an alkylating agent that when injected into male mice, induces mutations at a frequency of one per locus in every 700 gametes (Davis et al 1998). Clapcote and Roder screened the RIKEN ENU-based gene-driven mutagenesis system (RGDMS) for *Disc1* mutants and found to two independent point mutations, Q31L and L100P, in exon 2 of the mouse *Disc1* gene (figure 1.1) in C57BL6/J sperm, which was then used to impregnate DBA female mice. First generation individuals carrying either mutation were backcrossed with C57BL6/J mice for 6 generations until a predominantly C57BL6/J background line (average 98.5% at N6) was intercrossed to generate homozygous and heterozygous mutation carriers.

The L100P and Q31L point mutations change the amino acid sequence and are predicted to alter the structure and therefore the function of the protein. The evidence suggests that in adult mice the gene is expressed at the same level as in wild types but that it may have

different functional properties. In the Q31L mutant the mutation from glutamine, a hydrophilic amino acid normally expected to be on the outer surface of the protein, to leucine, a hydrophobic amino acid expected to be found on the protein inner surface would be expected to cause a distinct conformation change. The Q31L mutation also disrupts a known pde4b binding site in the disc1 head domain. Cell line studies have shown that binding of pde4b is significantly reduced in both the *Disc1* mutated lines. In Q31L mice, the amount of pde4b protein is not altered from the wild type level, but activity is reduced by up to 50% [197]. The change from leucine to proline in the L100P line would also be expected to result in distinct structural alterations of the disc1 protein due to the ring structure of proline, which is known to cause a sharp transition in polypeptide chain direction. It is highly likely both mutations result in altered *Disc1* proteins at the structural level.

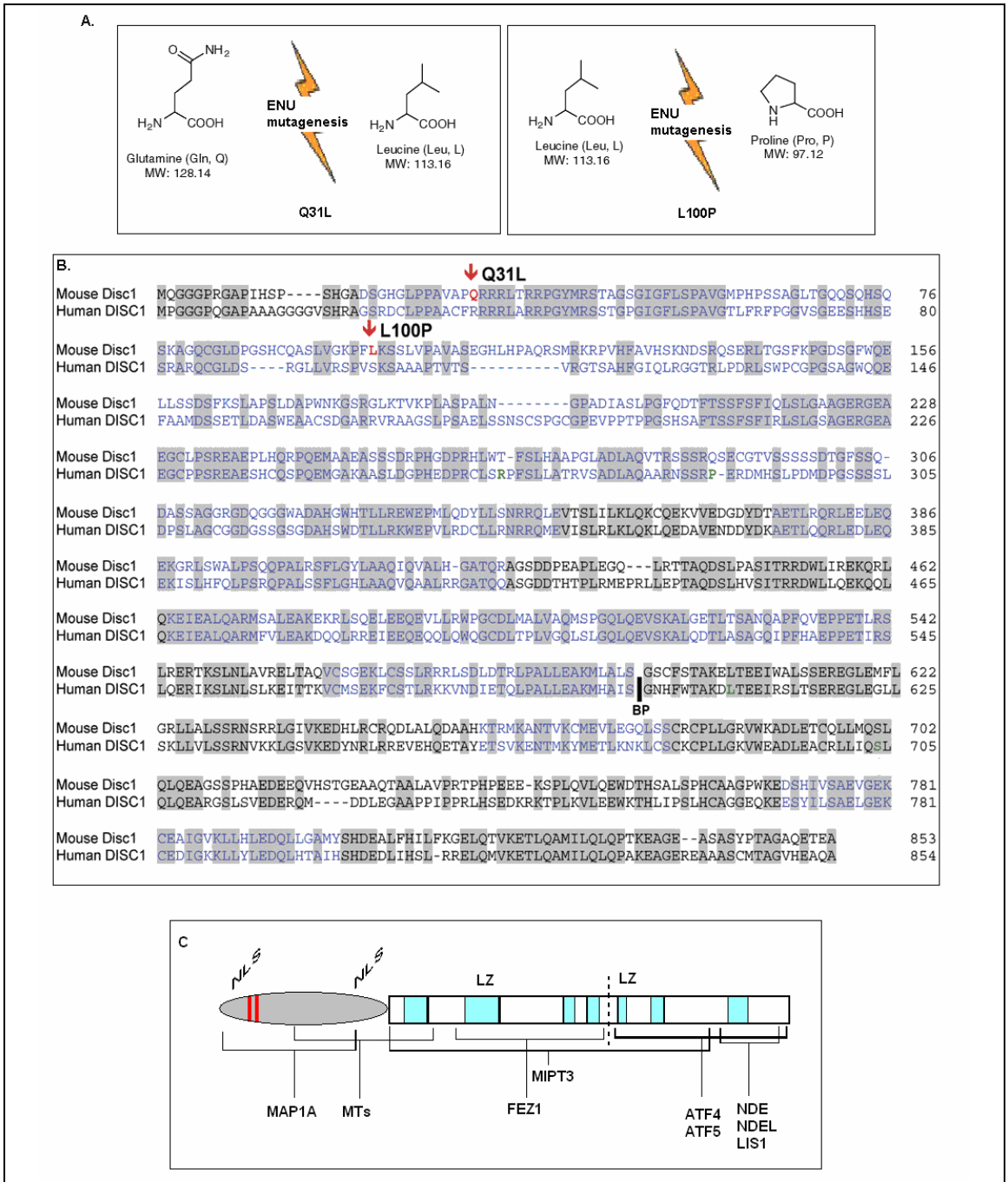

Figure 1.1 : The Disc1 Protein. (A) Amino acid changes in *Disc1* resulting from ENU mutatgenesis of C57B6/J sperm cells (B) Adapted from Clapcote *et al* 2007. Amino acid sequence of the Disc1 protein with the two ENU mutations highlighted. Conserved amino acids between mouse and human are shaded in grey. (C) Diagram of the Disc1 protein with known regions of interaction. The area directly affected by the ENU mutations is shown by two red lines in the head domain. *NLS* = nuclear localisation sequence, *LZ* = leucine zipper.

Prepulse inhibition and latent inhibition are common methods used to quantify information processing (sensory-motor gating) deficits in schizophrenia [107]. Mice homozygous for L100P and Q31L had lower prepulse inhibition (degree to which the startle response is reduced when startle eliciting stimulus is preceded by non-startle eliciting stimulus) than wild type controls, and latent inhibition (the degree to which exposure to a conditioned stimulus decreases the salience of a paired unconditioned stimulus) was impaired in both lines. Behaviours associated with depression were observed using the forced swim test, social interaction test and reward response test in Q31L mice only (Table 1.2). Homozygote Q31L mice displayed high immobility in forced swim tests and avoided social interaction and reward responsiveness [197].

	Change compared to wild type	
Behaviour	Q31L hom	L100P hom
Anxiety	=	=
Startle reactivity	=	reduced
Spatial learning and memory	=	=
Taste Sensitivity	=	n/a
Olfactory Function	=	=
Prepulse Inhibition	reduced	greatly reduced
Latent inhibition	greatly reduced	greatly reduced
Social interaction	reduced	=
Reward Responsiveness	reduced	=
Immobility during forced swim test	increased	=

Table 1.2: Common behavioural traits and the effects of the ENU missense mutation on phenotype when compared to the C57B6/J parent line = no significant change. Adapted from Clapcote *et al* 2007

Widely used antipsychotic and mood stabilising drugs were shown to reverse these behavioural anomalies, almost to the levels of non drug treated wild type controls (Table 3). Interestingly, treatment with the antipsychotics clozapine and haloperidol was effective in correcting the behavioural abnormalities in the L100P mutants, as was the PDE4 inhibitor rolipram, while these treatments had no effect on correcting the majority of the deficits observed in the Q31L line. Conversely the antidepressant biogenic amine

reuptake inhibitor Bupropion was effective in rescuing behaviour in Q31L but not L100P mice [197].

Drug Treatment		Q31L hom	L100P hom
Pre-pulse inhibition	Clozapine	=	positive p<0.05
	Haloperidol	=	positive p<0.05
	Rolipram	=	positive p<0.001
	Bupropion	positive p<0.001	=
Latent inhibition	Clozapine	=	positive p<0.05
Horizontal Activity	Clozapine	positive p<0.001	positive p<0.01
Forced swim test	Bupropion	positive p<0.01	n/a
	Rolipram	=	n/a

Table 1.3: The effects of drug treatment on behavioural rescue in Q31L and L100P missense mutant mouse lines. Adapted from Clapcote *et al* 2007

From the pharmacological responses and distinct behavioural phenotypes observed Clapcote *et al* [197] categorized the Q31L mutation as depressive-like and the L100P mutation as schizophrenia-like. They also observed reductions in brain size coupled with tissue shrinkage in cortex, thalamus and cerebellum. These areas correspond well to known neuroanatomical features of schizophrenia. This could be due to the predicted altered function of DISC1 in these mutants during brain development, but prior to this time a developmental profile of whole genome gene expression has not been obtained for the ENU missense mutant lines.

1.8 Thesis Aims

There were four main aims of this thesis.

1. to determine the effect of the two missense mutations on whole genome gene expression through development and into adulthood
2. to determine the effects of certain drug treatments on gene expression in adult mice ultimately allowing us to determine potential target genes, or pathways for targeting by drug treatment.
3. to determine gene expression levels throughout embryonic and early post-natal development of genes identified in the initial study.
4. to determine the effects of the *disc1* mutations on protein expression in the adult mouse and mature neuronal cultures

.

My hypothesis was that the expression levels of genes involved in brain function and development would be altered in the mutant mouse lines. I proposed that these alterations would occur from early embryonic stages, correlating with previous reports of delays in neuronal migration and maturation in *Disc1* mouse models. My secondary hypothesis would suggest that administration of drugs previously shown to correct aberrant patterns of behaviour in this mouse model may correct the expression patterns of these genes in the tissues tested. This would correlate with the changes in endophenotypes and patterns of behaviour observed in previous studies. It would be hypothesized that if this is the case it is a secondary effect (due to the action of the drug binding to receptors and preventing neurotransmitter release) and not the primary action of the drug, but may give some indication of the drugs mode of action and future development potential.

Chapter 2
Materials and Methods

2.1 Collection of tissue samples

In order to generate a wild type *Disc1* developmental profile, C57BL6 female mice were obtained from Harlan, UK and bred to males of the same strain held in-house at the Western General Hospital Biomedical Research Unit. Females were taken at key time points throughout gestation based on previous results from Shurov *et al [134]*, and sacrificed by cervical dislocation. Pups taken for post-natal developmental stages were sacrificed by decapitation (if under 2weeks) or cervical dislocation. C57BL/6 is an inbred laboratory mouse strain widely used in research as a wild type or background strain as it has a permissive background for maximal expression of most mutations (www.jax.org). They are long lived and breed well.

Missense mutant mice for the microarray study were obtained from Steve Clapcote and John Roder at the Samuel Lunenfeld Research Institute, Toronto, Canada. Wild type C57BL6J mice were obtained from Jackson Laboratories (www.jax.org). Where possible, matings were set up between homozygous individuals to remove the need to genotype all animals. Adults of 12 weeks used for the drug treatment trial and developmental stages were obtained as above.

2.1.1 Use of Inbred Strains

The C57BL/6J inbred strain was created by Dr. CC Little (who went on to found the Jackson laboratories) in the 1950's and all mice in this strain are derived from this original mating. All C57BL6 mice carry a mutation $Cdh23^{753A}$ which causes a frame skip of exon 7 in Cadherin 23 resulting in age related hearing loss at around 10months of age. Subsequent splitting of the original C57BL6 line has resulted in the arrival of new spontaneous mutations specific to subpopulations. One such mutation results in the deletion of the alpha-synuclein locus (which has been implicated in Parkinsons disease) in C57BL6 mice from the subpopulation available from Harlen UK (www.harlan.com). While this is not thought to affect other genes in this population we wanted to ensure our background strain was as close to that of the missense mutants as possible.

The L100P and Q31L mutants were backcrossed for several generations with C57BL/6J mice from the Jackson Laboratories (which do not carry the alpha-synuclein deletion). Due to time restrictions and problems achieving successful matings while in Toronto it was deemed necessary to complete the collection of C57BL6 embryonic stages back in the UK. The animals immediately available to us were C57BL6 mice from the Harlan subpopulation, so to ensure continuity it was necessary to order new breeders from Charles River (www.criver.com), the only UK supplier of the Jackson C57BL/6J substrain. This done, it meant we could confidently make direct comparison to those samples collected in Toronto. It would however be interesting to compare gene expression between the Harlan and Jackson subpopulations to determine if the spontaneous mutation would have had an effect on the experimental outcome.

All necessary measures were taken to ensure minimal stress to animals prior to sacrifice.

2.1.2 Extraction of Genomic DNA from adult tissue

Mice were earmarked for identification and the ear notches used to extract DNA for genotyping. 300µl 50mM NaOH was added to the ear notches at room temperature and then incubated at 95°C for 15-20minutes on a heat block to break down the tissues. Samples were mixed by vortex and pulse spun in a centrifuge to collect any condensation. 50µl 1M Tris-HCl was added and samples again mixed by vortex for 15 seconds. Samples were spun in a centrifuge at 13,000rpm for 6 minutes and the supernatant collected for genotyping. Samples were stored at -20°C until required.

2.1.3 Extraction of Genomic DNA from embryonic tissue

Genomic DNA extraction was carried out on embryonic samples from E10 and E13 pups, and from those animals for which genotype had to be ascertained. Extractions were carried out using the DNeasy blood and tissue kit (Qiagen) as per the manufacturer's protocol. Tissue (either a leg for embryonic samples or a tail clipping for post natal samples) was cut into small pieces and placed in a 1.5ml eppendorff with 180µl buffer

ATL and 20µl proteinase K. The sample was mixed by vortexing and incubated at 56ºC for 4 hours to ensure complete lysis of the tissues. Samples were vortexed and 200µl buffer AL added. 200µl of 100% ethanol was added immediately and the sample mixed by vortexing. The mixture was transferred to a DNeasy mini column in a 2ml collection tube and centrifuged at 8000rpm for 1 minute. Flow through was discarded and 500µl of wash buffer AW1 added to the column. This was then centrifuged for 1 minute at 8000rpm and flow through was discarded and 500µl of wash buffer AW2 added to the column. This was centrifuged for 3 minutes at 12,000rpm to dry the spin column and flow through and collection tube discarded. The spin column was placed in a new 1.5ml eppendorff and 200µl buffer AE added directly to the membrane. The column was incubated for 1 minute at room temperature and then centrifuged for 1 minute at 10,000rpm to elute the DNA. DNA was stored at -20ºC until required.

2.1.4 Genotyping by sequencing reaction

2.1.4.1 Polymerase Chain Reaction (PCR)

A master mix was prepared on ice consisting of 1X Sigma PCR buffer (SIGMA), 0.1mM of each forward and reverse primer (Invitrogen), 0.6µM dNTPs and 1U Sigma *Taq* DNA polymerase (SIGMA). Typically 0.5µg gDNA was used for a 20µl reaction volume and the master mix aliquoted into the tubes. PCR reactions were run on the Peltier Thermal Cycler-225 (MJ Research) and heated lids used to minimise evaporation. The initial PCR was run under cycle conditions; 94ºC 10mins, 94ºC 1min, 55ºC 1min, 72ºC 1min (return to stage 2 and repeat 30times), 72ºC 5mins, hold at 4ºC. Samples were run on a 1.5% agarose gel to check for product.

2.1.4.2 Exosapit Clean-up

2µl product from the PCR was added to 2µl H2O and 1µl Exosapit for a 5µl reaction mix. For each sample there were two reactions, to allow the use of forward and reverse primers in the next stage. The plate was run on the Peltier Thermal Cycler-225 (MJ Research) at 37ºC 1hour, 80ºC 20mins and held at 4ºC.

2.1.4.3 Sequencing reaction

A master mix was prepared consisting of 0.5µl BD v.3.1, 1.75µl Sequencing buffer, 1µl 3.2pmol Primer, 2.75µl H2O to a total volume of 6µlper sample. This was added to each sample in the plate used for the previous reactions and run on the Peltier Thermal Cycler-225 (MJ Research) under cycle conditions; 96ºC 1min, 96ºC 10secs, 50ºC 4secs, 60ºC 4mins (return to stage 2 and repeat 30times), hold at 4ºC. After PCR the samples were stored at -20ºC for ethanol/EDTA precipitation the following day.

2.1.4.4 Ethanol/EDTA Precipitation

Sample plate was allowed to equilibrate to room temperature and then pulse spun in a centrifuge to collect any condensation. 2.5µl 125mM EDTA was added to each well along with 30µl 100% EtOH and hybaid sealing mat replaced. Reactions were mixed by inversion 4times and incubated at room temperature for 15mins. Plates were spun in a Jouan centrifuge at 3000rpm for 30mins and seal removed before plate was inverted over a paper towel to remove most of the EtOH. The paper towel was placed in the bottom of a centrifuge bucket and plate spun inverted up to 1000rpm then stopped. 40µl freshly diluted 70% EtOH was added to each well and the seal replaced before mixing by inversion 4times. The plate was spun in Jouan centrifuge at 3000rpm for 15mins and seal removed before plate was inverted over a paper towel to remove most of the EtOH. The paper towel was placed in the bottom of a centrifuge bucket and plate spun inverted for 30secs then stopped. Wells were air dried for 15mins protected from light and the plate sealed with adhesive film before being sent to the Medical Research Council sequencing service to be sequenced by Agnes Gall.

2.1.5 Genotyping using the Transnetyx sequencing service

Ear notches were collected from mice and placed in the 96-well plate provided by Transnetyx services (www.transnetyx.com). Positive and Negative control samples from mice with known genotypes were included in the first run. Samples were sent by courier to Transnetyx to be analyzed using primers designed from sequence submitted by myself. Results were returned within 3 days.

2.1.6 Genotyping Q31L het x hom pups

Genotyping of Q31L pups from het x hom matings was carried out by Edward Weiss at the Toronto Sick Kids Hospital Research laboratories.

2.1.7 Dissection of embryos and pups

Pregnant females were sacrificed by cervical dislocation and observed until all movement had ceased. An incision was made into the females abdominal cavity and the skin pulled back to reveal the embryonic sacs. Embryos were removed and placed in chilled PBS on ice. Under a standard dissecting microscope the heads were removed and the top of the skull removed to expose the brain. The brain was extracted whole and placed in prechilled RNA*later* (Ambion) at 4ºC to stabilize the RNA for further analysis, and kept at 4ºC for up to 2 weeks before required. For post-natal samples the heads were also removed and brains extracted in the same manner as the embryonic stages.

2.1.8 Dissection of adult brain samples

Adults were decapitated post mortem and the skin cut from the back of the head to the nose down the midline. The tissue was then peeled back to reveal the top of the skull. This was removed by making an incision around the circumference of the skull starting at the foramen magnum and proceeding round the top of the skull at the level of the occipital openings. The top of the skull was lifted off the brain and the whole brain removed intact from the skull and placed dorsal side up on filter paper moistened with PBS. The cerebellum was removed by making a coronal cut perpendicular to the paper just behind the inferior colliculi, and placed in pre-chilled RNA*later* (Ambion). A sagittal incision was made down the midline between the two hemispheres to one third from the anterior end. Using two paintbrushes the hemispheres were separated exposing the hippocampal groove. One brush was slid into the grove, separating the hippocampus from the cortex, and the hippocampus gently rolled out. Once separated from the rest of the brain the hippocampus was placed in prechilled RNA*later* prior to further analysis. This was repeated for the other hemisphere also. The cortex was gently peeled away and

stored in RNA*later*. In some cases only the hippocampus was removed in which case the cortex and remaining brain tissues were stored in the same eppendorf as the cerebellum. Where no further dissection was being carried out the brain was removed from the skull whole and placed directly in pre-chilled RNA*later*.

2.2 Sample Preparation

2.2.1 Extraction of total RNA from tissue samples

Total RNA was isolated from selected tissue in accordance with the protocol supplied with the RNeasy mini kit for animal tissues (Qiagen). Tissues were homogenized in 600µl buffer RLT using a Precellys 24 tissue homogenizer (Bertin Technologies) at maximum speed for two 45 second cycles. Lysate was centrifuged for 3 minutes at 13,000rpm and the supernatant transferred to a fresh micro centrifuge tube. 1 volume of 70% ethanol was added to the lysate and mixed by pipetting and 700µl of this sample transferred to an RNeasy spin column. After centrifugation at 10,000rpm for 15seconds, flow through was retained for protein analysis and 700µl buffer RW1 added to the spin column. Columns were again centrifuged under the same conditions and flow through retained and combined with that from the previous step. The spin column membrane was washed by addition of 500µl buffer RPE and spun for 15seconds at 10,000rpm. Flow through was again retained and the step repeated with a 2 minute spin time to dry the membrane. RNA was eluted by the application of 40 µl sterile H_2O directly to the membrane and centrifuged for 1 minute at 10,000rpm. Pooled flow through was stored at -20°C until required for protein analysis. Concentration and quality of total RNA samples (RIN) was determined using a spectrum analyzer (Agilent Bioanalyser ; see section 2.2.5) and samples stored at -20°C until required for cDNA synthesis.

The same method was followed for RNA extraction of the samples collected at the Samuel Lunenfeld Research Institute however due to lack of access to the Precellys 24 tissue homogenizer, all tissues were homogenized using a standard mortar and pestle. These samples also underwent on column DNA digest (see section 2.2.3).

2.2.2 PARIS procedure

The PARIS procedure from Qiagen allows simultaneous extraction of RNA and protein from the same sample tissue. This enables us to use the same animal and sample section for subsequent protein analysis after the initial RNA/DNA analysis is complete. The PARIS procedure was used on animals for the second round qRT-PCR Taqman analysis.

Lysate is prepared by placing tissue in a homogenizer tube with 600µl ice-cold Cell disruption buffer (provided with kit) and homogenizing in a rotor homogenizer for 30secs. The lysate was split into two eppendorf tubes of equal volume. One tube was stored at -80°C for later protein analysis while 300µl 2X lysis/binding solution was added to the other tube and mixed by pipetting. One volume 100% EtOH was added and the sample was mixed by pipetting to ensure the EtOH reached the bottom of the tube. The sample mix was transferred into a filter cartridge and spun at 10,000rpm for 1min in a centrifuge. Flow-through was discarded and 350µl was solution 1 added to the cartridge. The sample was spun at 10,000rpm for 30secs and flow-through again discarded. 500µl wash solution 2/3 was added to the cartridge and the sample was spun at 10,000 rpm for 1min. Flow-through was discarded and this step repeated a second time. The empty cartridge was spun at 10,000rpm for 15secs to remove any residual wash solution and transferred to a fresh collection tube. 40µl pre-heated (to 95ºC) elution solution was added to the cartridge membrane and the sample was spun at 10,000rpm for 30secs. This was repeated so a total of 80µl RNA was collected. Any genomic contamination was removed using the DNA-*free* kit (Ambion) outlined in section 2.2.3.

2.2.3 Post extraction DNase Treatment of RNA

RNA samples were treated with the DNA-*free* kit (Ambion) prior to cDNA synthesis to remove any traces of genomic contaminant. Treatment was carried out as per the protocol supplied with the kit. 0.1 volumes of 10x DNase1 buffer and 1µl rDNase 1 was added to the samples and mixed gently before incubation in a 37°C water bath for 30minutes. 0.1 volume of DNase inactivation reagent was added to the samples and incubated at room temperature for 2 minutes ensuring the samples were continually mixed over the incubation period to re-disperse the inactivation reagent. After centrifugation at

13,000rpm for 2 minutes the supernatant containing the RNA was transferred to a fresh tube to be used for cDNA synthesis.

2.2.4 On column DNase Treatment of RNA

On column DNA digest was carried after the preliminary wash stage of RNA extraction using the RNase-Free DNase set (Qiagen) by adding 70µl Buffer RDD to 10µl DNase 1 and applying this mix directly to the spin column membrane. The column was then incubated at room temperature for 15-20 minutes to allow complete DNA digest to occur. The membrane was washed again with 350µl buffer RW1 and RNA extraction continued.

2.2.5 Assessing RNA quality using the Agilent Bioanalyser

The Agilent 2100 Bioanalyser uses electrophoresis to separate fragments of the RNA sample by size. Quantification of each fragment allows quality control to be carried out at the same time as checking quantity. Samples collected at the Samuel Lunenfeld Research Institute in Toronto were analyzed by me on my return to Edinburgh. All other samples, including those collected in-house at the Molecular Medicine Centre and those which required vacuum drying prior to amplification and labeling for the microarray study, were analyzed by Alison Condie at the Welcome Trust Clinical Research Facility, Western General Hospital, Edinburgh.

The RNA 6000 Nano kit was placed on the bench for 30mins prior to loading the chip to allow the dye concentrate to equilibrate to room temperature. 550µl of RNA 6000Gel matrix was pipetted into a spin column and centrifuged for 10 minutes at 1500g. Gel was split into 65µl aliquots to be used within 4 weeks. Once at room temperature the RNA 6000Nano dye concentrate was placed on a 10second pulse spin and 1µl added to a 65µl aliquot of filtered gel. The solution was vortexed for 15seconds to mix and centrifuged for 10mins at 13000g. RNA samples were denatured for 2minutes at 72ºC and placed on ice ready for loading onto the chip.

The RNA 6000 Nano chip was placed in the chip priming station and 9µl gel-dye mix added to the well marked  and the priming station closed. The plunger was pressed until held in the clip and released after 30 seconds. After 5 seconds the plunger was pulled back to the 1ml position and the station opened. 9µl of gel-dye mix was pipetted into the wells marked  and the remaining gel mix discarded. The RNA 6000 Nano marker was loaded into all 12 sample wells and the ladder well. 1µl of the RNA 6000 Nano ladder was added to the appropriated well marked  and 1µl of sample into each of the 12 sample wells. The chip was placed horizontally in the IKA vortex and vortexed for 1 minute at 2400rpm. The chip was placed in the Agilent 2100 Bioanalyser and run within 5 minutes of vortexing. The programme selected quantifies total RNA in the sample.

After the bioanalyser had completed its programme the chip was discarded and the bioanalyser pins cleaned by immersing in 500µl RNA zap for 2mins then 500µl distilled H_2O for 2 mins before being air dried for 10 seconds. Total RNA results were analyzed using the Agilent 2100Expert software. Samples with a RNA Integrity Number (RIN) of <5 were deemed unsuitable for use in the microarray study due to problems with amplifying low quality RNA (Figure 2.1).

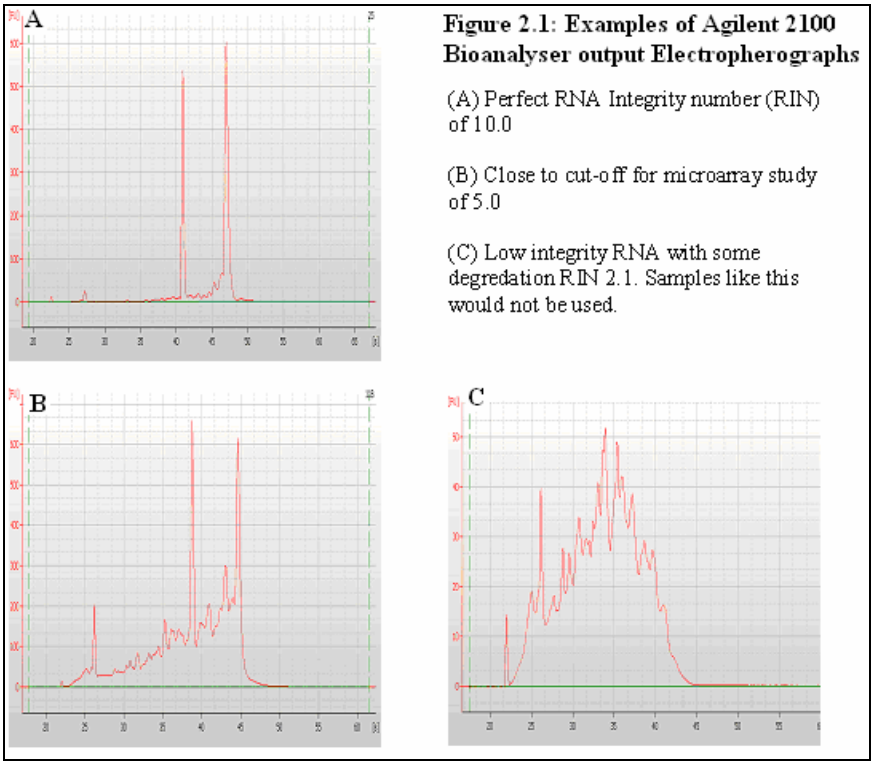

Figure 2.1: Examples of Agilent 2100 Bioanalyser output Electropherographs

(A) Perfect RNA Integrity number (RIN) of 10.0

(B) Close to cut-off for microarray study of 5.0

(C) Low integrity RNA with some degredation RIN 2.1. Samples like this would not be used.

Of all samples collected (both in house at the MMC and at the Samuel Lunenfeld Research Institute) 96.91% met the RIN number criteria for progressing forward for the microarray. Of those groups chosen for the microarray study 95.10% had suitable RNA for amplification.

2.2.6 Vacuum Drying RNA samples

While the majority of samples (97%) collected for the microarray study contained enough total RNA to go forward for amplification and labeling, a few had very low concentrations and would require drying before amplification. This was achieved using a vacuum centrifuge.

The centrifuge was cleaned with RNAzap to prevent cross contamination and was set to heat to 60ºC for 10minutes prior to loading the samples. Samples were loaded in 1.5ml eppendorffs with the lids removed and the vacuum pump and centrifuge set running. Progress was checked every 5minutes to ensure the samples were not over-dried and the

running time was adjusted according to the volume left in the tubes at any time throughout the process. All samples were dried from 45µl down to around 15µl over a period of around 35 minutes. An aliquot of the concentrated sample was stored for quantification on the Agilent Bioanalyser and the remaining sample stored at -70ºC until required for amplification. Some samples returned poor RIN numbers post drying and so could not be used for the microarray study. The other samples, with RIN's above 5, were deemed acceptable for amplification and would be checked after this to assess suitability.

2.2.7 Synthesis of cDNA

Single-stranded cDNA was synthesized from DNase treated total RNA using the core cDNA synthesis kit (Roche). 2µg total RNA was diluted in sterile H_2O to a total volume of 16.4µl and denatured at 65°C for 15minutes before being placed on ice for 5minutes. This was then incubated at 47°C for 60minutes in the presence of 10x buffer, 25mM MgCl2, 5mM dNTPs, OligoT, RNase inhibitor and AMV reverse transcriptase. Controls underwent the same conditions in the absence of AMV reverse transcriptase or RNA. The reaction was terminated by heat inactivation at 99°C for 5 minutes. Single stranded cDNA was stored at -20°C until required for PCR amplification.

2.2.8 Synthesis of cDNA from low quantity RNA

Synthesis of single stranded cDNA from low quantity RNA was carried out using the Transcriptor cDNA synthesis kit from Roche. 2µg total RNA was diluted in sterile H_2O to a total volume of 22µl and added to 4µl random primers. The mix was denatured at 65°C for 15minutes before being placed on ice for 5minutes. This was then run of a cycle of 25°C for 10 minutes, 55°C for 30 minutes and 85°C for 10 minutes in the presence of 5x buffer, 5mM dNTPs, RNase inhibitor and AMV reverse transcriptase. After completion of the cycle the samples were held on ice for 5 minutes and transferred to -20°C until required for further experiments.

2.2.9 Determining cDNA concentration using the Quant-iT™ PicoGreen® dsDNA kit

Concentration of cDNA preparations was determined using the Quant-iT™ PicoGreen® dsDNA kit (Invitrogen). The kit contains the Quant-iT™ PicoGreen® dsDNA reagent which stains double stranded DNA in solution and fluoresces when excited at 480nm in a Cytofluor Fluorimeter (at the Cancer Research Facility, Western General Hospital, Edinburgh). Intensity of fluorescence emission of samples and standards was measured at 520nm.

The 100µg/ml lambda DNA standard supplied with the kit was diluted 50 fold in TE to give a 2µg/ml DNA stock for preparing standards. Standards were prepared at 0, 1, 10, 100 and 1000ng/ml DNA and DNA samples to be measured were diluted in duplicate in TE (5µl sample + 95µl TE) in a DynaTech 96 well plate. As the predicted yield of the cDNA reaction is 30% all samples should fall within the range of the standard curve.
The PicoGreen reagent was diluted 200 fold in TE and covered in foil to protect from light. The reagent must be used within a few hours to be most reliable so only enough working solution for one run was prepared allowing 100µl for each sample and standard. The fluorimeter was set to excite at 480nm and measure emission at 520nm with a gain of 45 so the highest DNA standard would be close to the fluorimeter's maximum. The Picogreen working solution was added to the samples and readings taken immediately. Emission readings of the DNA standards were plotted and concentration of cDNA samples calculated using the regression equation of the standard curve.

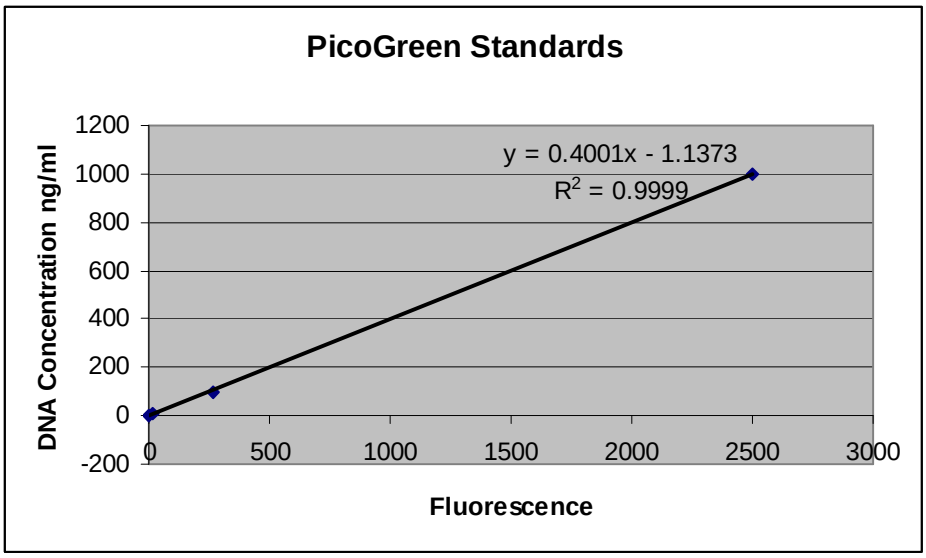

Figure 2.2: Picogreen standard curve for assessing cDNA concentration.

2.2.10 Determining cDNA or RNA quantity on the NanoDrop

The Thermo Scientific NanoDrop 1000 spectrophotometer is a highly accurate full spectrum spectrophotometer for measuring concentration and quality of nucleic acids and proteins. As little as 1µl of sample can be used to determine concentrations, reducing the amount of sample required compared to traditional methods. Sample concentration is determined using the Beer-Lambert equation (A= e*b*c) where **A** is the sample absorbance measurement (-log[sample intensity/blank intensity]), **e** is the wavelength-dependent extinction coefficient (liter/mol-cm), **b** is the path length in centimeters, and **c** is the analyte molarity (M). On start up the nucleic acid option was chosen and for cDNA the 'ssDNA33' programme was selected, and for RNA the 'RNA-40' programme was selected. Selection of the correct programme is essential for accurate calculation of concentrations.

The machine was blanked using 1µl dH_2O and this is stored as reference for calculating the subsequent samples. Each sample was tested on the NanoDrop ensuring the pedestal was cleaned after every sample. Concentrations were saved in the report and absorbance ratios and curves checked for quality. The 260/280 ratio gives a comparative value of

absorbance at 260 and 280nm and is used to assess quality of RNA and DNA. A 260/280 ratio of 1.8 is accepted as 'pure' for DNA and a ratio of 2.0 'pure' for RNA. Deviations from these values would suggest contamination of the sample. Any samples run which were not sufficiently close to these ratios were discarded from further study.

2.3 PCR

Polymerase chain reactions (PCR) allow the rapid amplification of a selected region of DNA sequence from a complex mixture. Specificity is provided by oligonucleotide primers designed to be complimentary to the 5´ region of sequence to be amplified. The PCR reaction is a three stage process. First the DNA is denatured at a high temperature, and then the temperature is reduced to allowing annealing of the oligonucleotides to the specific DNA sequence. Finally the temperature is adjusted again to allow elongation and amplification by a thermo-stable DNA polymerase. Multiple cycling of these stages results in exponential amplification of the desired sequence.

2.3.1 PCR Reagents

As only RNA expressed at certain time points was of interest, reverse transcriptase PCR (RT-PCR) was used. A master mix was prepared on ice consisting of 1X Sigma PCR buffer (SIGMA), 0.1mM of each forward and reverse primer (Invitrogen), 0.6µM dNTPs and 1U Sigma *Taq* DNA polymerase (SIGMA). Typically 0.1µg cDNA was used for a 20µl reaction volume and the master mix aliquoted into the tubes. PCR reactions were run on the Peltier Thermal Cycler-225 (MJ Research) and heated lids used to minimise evaporation.

2.3.2 Design and Synthesis of Oligonucleotide Primer pairs

Oligonucleotides were designed using the *Primer3* software (primer3_www.cgi v 0.2) with in-house default features. All oligonucleotides were 18-25bp long with an average

50% GC content and a melting temperature (T_m) of 60-65ºC. No primers chosen were predicted to dimerise or hairpin when used as a pair. All primer pairs were commercially sythesised by Invitrogen at 25nM scale with standard purification.

2.3.3 RT-PCR Cycling

The PCR consisted of an initial denaturing step of 94ºC for 2 mins followed by 30 cycles of denaturation, annealing and extension. The denaturation step lasted 30 seconds at 94ºC. The annealing temperature was determined by the predicted Tm of the oligonucleotide primers. Primers were typically designed with a Tm of 64°C so the annealing step was carried out 5ºC below this at 59ºC for 30seconds. The extension step was carried out at 72ºC, the optimum temperature for *Taq* DNA polymerase, and the length of stage determined by final predicted product size. Typically 1 minute per 1Kb of sequence was used. After the final cycle a further elongation step of 94ºC for 5minutes was used to ensure fragments were complete.

The 'Touch-down' PCR method was used as standard as it increases specificity of the PCR reaction. The initial annealing temperature used is 10 ºC above the final annealing temperature required, and drops by 1ºC with every cycle until after ten cycles the final annealing temperature is reached. The normal 30cycle reaction at the desired annealing temperature can than proceed. This method results in highly specific primer annealing in the first few cycles selecting for the correct product in the subsequent cycling sets.

2.3.4 Visualisation of PCR Products

PCR products were separated by electrophoreses on agarose gel of between 1 and 2% (w/v) depending on the size of the expected bands (gel used is noted on figure legends). Multi-purpose agarose was boiled in 0.5x TBE buffer (45 mM Tris base, 45 mM boric acid, 1 mM EDTA pH 8.0) and SYBRsafe (Invitrogen) added 1:50 (v/v). Gel was left to set in the dark to prevent any bleaching of the SYBR safe. 2µl loading dye was added to the PCR product prior to loading on the gel and 400ng of 1Kb ladder (Invitrogen) run

alongside the samples. Products were visualized on a UV transilluminator and photographed with a digital camera.

2.4 Protein Preparations

2.4.1 Determining Protein Sample Concentrations

Concentrations of protein samples obtained during the PARIS procedure were determined using the BCA Protein Assay Kit (Pierce Biotechnologies) test tube procedure as per the manufacturers' instructions. This relies on the reduction of copper ions by protein in an alkaline medium which is detected by kit reagent A. Protein standards of known concentrations from 0-2000µg were initially prepared by dilution of 2mg/ml stock Albumin. BCA working reagent was prepared by mixing 50 parts BCA reagent A with 1 part BCA reagent B. Total volume required was calculated as:

(#standards + #unknowns) x (#replicates) x 2 = Total volume working reagent required

0.1ml of each standard and unknown sample replicate was pipetted into an appropriately labeled tube and 2ml working reagent added to each. Tubes were incubated at 37ºC in a waterbath for 30 minutes and then allowed to cool to room temperature. As the BCA reaction is not an end-point colour change, the colour will continue to develop after cooling to room temperature. However, the rate of colour change is so low at room temperature that no significant error will be recorded if all samples are measured within 10 minutes of each other. Sample absorbance was measured at 562nm on a spectrophotometer. The absorbance reading for the blank sample was subtracted from subsequent standard and unknown sample measurements. These blank correlated BSA standards were plotted on a standard curve against their know concentrations and this was used to determine the protein concentrations of the unknown samples.

2.5 Pre-Pulse Inhibition

Pre-pulse inhibition is a common method used to quantify information processing deficits in schizophrenia with reasonable validity. It can be used with both animal and human experiments. PPI measures the degree to which the acoustic startle response is reduced when a startle eliciting stimulus is preceded by a brief low-intensity stimulus which does not elicit a startle response.

2.5.1 Calibration

Animals from 11.5-14 weeks of age were weighed and the chamber platform of the prepulse box was calibrated to take into account the weight of the animal.

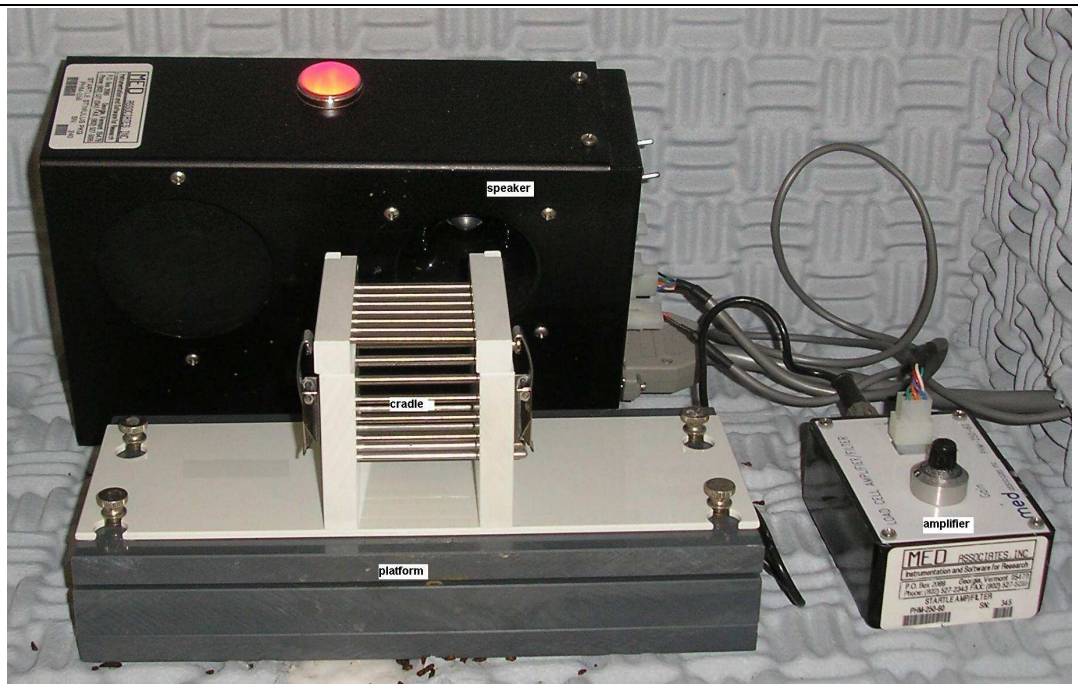

Figure 2.3: Image of a PrePulse inhibition chamber. Platform, cradle, amplifier and speaker are labeled.

2.5.2 Drug Administration

Drugs were administered 30 minutes before the commencement of behavioural testing by intra-peritoneal injection in a volume of 10ml/kg by Tatiana Lipina at the Samuel Lunenfeld Research Institute. Clozapine (3mg/kg; Tocris) was dissolved in saline (0.9% NaCl) containing 0.3% Tween (Biorad). Bupropion (4mg/kg; Sigma) was dissolved in distilled water and Rolipram (0.5mg/kg; Sigma) in saline containing 10% DMSO. Saline was used as a control for drug naïve animals. All drug doses were selected from previous studies carried out at this lab (Clapcote *et al* 2007 [197]).

Each genotype was split into groups for drug treatment. For the L100P mutant line and the C57BL6 line there were 12 mice in each drug treatment group of equal numbers male and female, including a control group administered with saline. For the Q31L mutant line 8 mice were used for each drug treatment with equal numbers of male and females. Rolipram was not administered to the Q31L line as it had been deemed unsuccessful at behavioural rescue in previous studies [197]. In all cases animals were sacrificed 1 hour after drug administration.

2.5.3 Behavioural Testing

Animals were placed in a pre-pulse chamber 15 minutes after administration of drug treatment. The pre-pulse program begins with a 15 minute habituation period so actual testing did not begin until 30 minutes post drug administration. Animals are then played a series of five ‘startles’, and movement is detected by the platform at the base of the chamber. This is followed by 50 trials of startle, startle preceded by pre-pulse or no stimulus. The final stage is a further series of 5 pure startle stimuli.

Animals were sacrificed immediately at the end of the pre-pulse testing by cervical dislocation, and whole brain was extracted and stored in RNAlater before further dissection.

2.5.4 Analysis of Behavioural Testing Results

Analysis of the amplitude of acoustic startle response in the Pre-pulse inhibition trials was carried out in Microsoft Excel and Statistica for Windows analysis packages. Repeated measures analysis of variance was used to determine any significant differences between groups with appropriate post-hoc tests applied and a standard p-value cut-off or 0.05 applied.

2.6 Microarray Analysis of ENU mutant gene expression

This section describes the methods used to carry out the microarray study from amplification onward. Preparation and extraction of RNA was as described in section 2.2.

2.6.1 cRNA amplification and Biotinylation

RNA samples were gathered as previously described in sections 2.2.1-2.2.6. Biotinylated, amplified cRNA was generated using the Illumina® TotalPrep RNA amplification kit as per the manufacturers instructions.

100ng total RNA was used as starting material and master mixes were prepared as recommended using the Illumina® TotalPrep™ RNA Amplification Master Mix Calculator available online (www.ambion.com/techlab/mm_calcs/illumina_rna_totalprep_amp_calc.php). The two 2-hour incubation cycles (II.C and II.D) were performed in a Peltier Thermal Cycler and hybridisation (step II.F) was carried out with overnight incubation for 14 hours at 37ºC in a hybridisation chamber. The eppendorf's were wrapped in parafilm for this step to prevent loss of material through condensation. 1.5µl of the final elute, containing amplified, biotinylated cRNA, was set aside for quantification and quality control and the remainder stored at -70ºC until required.

2.6.2 Quantification and Quality control of cRNA

Samples were loaded onto the Agilent RNA nano chip as described in section 2.2.5. The chip was loaded into the bioanalyser and run under the 'RNA nano-cRNA' programme, designed to measure the quantity of cRNA in the sample. The read-out obtained from the bioanalyser gives a measure of concentration and contamination (as a percentage). Samples with high levels of contamination would not be amplified.

2.6.3 Preparation of Pooled cRNA

Due to the number of animals used for the study it was decided samples would be pooled before being hybridised to the array. Each pool contained 3 animals of the same sex from the same mouse line. In most cases there were 2 pools for each group and sex (table 2.1). After determining the concentration of each individual sample using the Agilent 2100 bioanalyser, equal amounts of each sample were combined to a total concentration of 150ng/µl.

Array	Beadchip						
	1	2	3	4	5	6	7
A	***C57BL6 Sal F-2***	***C57BL6 Sal F-2***	***C57BL6 Sal F-2***	***C57BL6 Sal F-2***	***C57BL6 Sal F-2***	***C57BL6 Sal F-2***	*C57BL6 Sal F1(2)*
B	C57BL6 Sal F-1	L100P Cloz F-1	Q31L E13 F-2	C57BL6 Sal M-2	L100P Roli F-2	L100P Sal F-2	*C57BL6 Sal F1(3)*
C	L100P Cloz F-2	L100P Sal F-1	C57BL6 Sal M-1	L100P Cloz M-2	C57BL6 E13 F-2	C57BL6 E13 M-2	*C57BL6 Sal F1(1)*
D	L100P Roli M-2	Q31L Sal F-1	L100P Cloz M-1	L100P Sal M-2	Q31L Bupro F-2	L100P Sal M-1	L100P E13 M-1
E	Q31L Bupro M-1	Q31L E13 M-1	Q31L Bupro F-1	Q31L Sal F-2	C57BL6 E13 F-1	L100P Roli F-1	C57BL6 E13 M-1
F	Q31L E13 M-2	Q31L Sal M-2	L100P Roli M-1	L100P E13 F-1	L100P E13 F-2	Q31L E13 F-1	Q31L Sal M-1

Key			
C57BL6	Wild type non-littermate control		
L100P	ENU mouse mutant with L100P Disc-1 mutation 'schizophrenic-like' phenotype		
Q31L	ENU mouse mutant with Q31L Disc-1 mutation 'depressive-like' phenotype		
Sal	Non-drug treated	E13	Embryonic day 13
Cloz	Clozapine treated	F#	Female pool #
Roli	Rolipram Treated	M#	Male pool #
Bupro	Bupropion treated		

Table 2.1: BeadChip layout with key. Numbers represent each individual beadchip and letters represent the position on the chip.

2.6.4 Hybridisation and scanning of Illumina BeadChips at the Wellcome Trust Clinical Research Facility

The Illumina MouseWG-6 v2.0 expression beadchip was used to target 45281genes from the RefSeq database release 22 and RIKEN FANTOM across six pooled samples per chip (http://www.illumina.com/downloads/GX_Mousev2.0_DataSheet.pdf). Each chip carried one identical sample to be used as a between chip control, and one sample set was amplified, pooled and hybridised 3 times to control for batch differences.

10 microlitres of each 150ng/µl sample was submitted to Alison Condie at the Genetics Core of the WTCRF where the subsequent hybridisation, washing, blocking and streptavadin-Cy3 staining was performed in line with manufacturers guidelines. The Illumina BeadArray™ Reader was used to scan the BeadChips and Illumina BeadScan software performed specific image processing steps to determine bead intensities from raw image data (as described in Kuhn *et al* 2004 [215]). Illumina BeadStudio 2.0 Gene Expression software was used to read the image data and generate signal data for subsequent analysis. Probe information including raw signal intensity, number of beads, standard deviation of raw bead intensity and detection *P*-value were exported to text file and returned to myself from the WTCRF. Two sets of data were initially returned, one with no normalisation and no grouping, and the other with no normalisation but with samples grouped according to line and treatment. This second sample set returned group detection *P*-values which would be used for data filtered later in analysis.

Unfortunately during the hybridisation stage one chip (chip 5) showed signs of drying out while in the chamber which resulted in some skewed intensity values of controls. It was decided that all samples on this chip should be re-run before further analysis. The chip 5 repeat gave control intensities comparable with previous runs and intensities from this were taken forward for further analysis. Results from both chip 5 runs are shown in Chapter 3.

2.6.5 Data Analysis

General descriptive statistics and many quality control methods were carried out using Microsoft® Office Excel with various add-ins, and Statistica for Windows. Venn diagrams were created using a combination of Microsoft® Office Excel and Microsoft Office Powerpoint. Most downstream analyses were performed using the free software environment R (www.r-project.org) and the R-based open source software for bioinformatics, Bioconductor (www.bioconductor.org).

2.6.6 Microarray Quality Control Measures

This section describes the methods used to analyse the inbuilt controls, within the Illumina BeadArray Platform, technical replicates and between chip controls outwith the standard chip controls. Cluster analysis has also been performed to assess the success for the experiment and quality of the data generated.

2.6.6.1 Illumina direct hybridisation assay controls

The Illumina BeadArray™ platform has six control categories inbuilt to assess the quality and reproducibility of sample intensities obtained as follows: 1) housekeeping controls assess the intactness of the biological sample; 2) negative controls provide an estimate of background noise; 3) hybridisation controls determine success of hybridisation; 4) biotin controls assess the success of biotin labelling and signal generation; 5) high stringency and 6) low stringency hybridisation controls assess the stringency of hybridisation. Controls 1,2 and 4 are dependant on the quality of the sample, while the others are not. It was at this stage the initial failing of the first chip 5 run was noticed and samples re-submitted.

2.6.6.2 Technical Replicates and Between Chip Controls

One sample (C57BL6Sal-F2) made up of 3 drug naive female C57BL6 mouse samples was loaded onto every chip analysed to assess BeadChip-BeadChip reproducibility. Another sample, prepared from the same total RNA but different cRNA preparations (C57BL6Sal-F1) was included in triplicate on one BeadChip as a measure of

experimental within chip variability. The reproducibility was assessed using the Pearsons correlation coefficient (r^2).

2.6.6.3 Cluster analysis for detection of experimental artifacts

Hierarchical cluster analysis was performed to determine whether samples clustered together through experimental variables such as batch effect or BeadChip. Various methods exist differing in measures of distance between pairs of samples and clusters. As this type of analysis had already been carried out by Andrea Christoforou within our group, the clustering method she used was followed as this had previously been shown to be suitable for similar sample sets. The R package **pvclust** (www.is.titech.ac.jp/~shimo/prog/pvclust/) was used to create a final dendogram and to assess the results via multisample bootstrap resampling. One thousand bootstraps were performed on each of three bootstrap sample sizes (r=0.5,1.0,1.5 where r is the size of the bootstrap sample relative to the actual dataset).

2.6.6.4 Power Calculations

Power calculations were performed using the *power.multi* function for multiple treatment designs from the Bioconductor **sizepower** package. No software package available at this time was able to produce power calculations for multiway ANOVA analysis so the calculations were carried out on the basis of 33 student t-tests on the smallest group size, giving a very conservative estimate of the experimental power. Curves were plotted for a range of standard deviations and effect sizes.

2.6.6.5 Pre-processing

Pre-processing is the steps taken to prepare data for analysis, and involved 3 stages: 1) variance stabilisation and transformation; 2) normalisation and 3) data filtering. The Bioconductor package **lumi** (www.bioconductor.org/packages/2.1/bioc/vignettes/lumi/inst/doc/lumi.pdf) which was created specifically for use with the Illumina BeadArray™ platform, was used for pre-processing. The first stage, variance stabilisation, aims to reduce the heteroskedacity (that is, make the variance more constant) across the sample set. This was carried out using the

lumi-vst function which is a generalised log transformation determined by direct modelling of the mean-variance relationship of bead intensity [216]. The second stage, normalisation, aims to remove or reduce systematic non-biological bias incurred by the experimental process. The *lumi-rsn* function was applied to the dataset for normalisation (www.bioconductor.org/packages/2.1/bioc/vignettes/lumi/inst/doc/Bioc2007_lumi_presentation.pdf).

The final stage of pre-processing was to filter the data. Data was filtered in two ways: First, any probes not expressed above background in at least one sample group were removed from further analysis. This was determined using the Illumina BeadStudio detection *P*-value, which is calculated for every probe in every sample using a non-parametric method that ranks the probe against that of a negative control. It is interpreted as the probability the observed signal is no greater than the background given the amount of noise in the data. Secondly, the probes were filtered based on fold change. Pairwise comparisons were made between samples grouped by strain and treatment and only probes with at least a 1.3 fold change (either above or below that of the reference group) were kept. The fold change cut off was chosen as this is the level of sensitivity offered by the Illumina platform (www.illumina.com/downloads/GX_Mousev2.0_DataSheet.pdf). Fold change was calculated by taking the inverse/anti-$\log_2$ of the difference between the average nomalised values. Cluster analysis was carried out after both filtering stages to check the effect on data distribution.

2.7 Testing For Differential Expression

This section describes the methods used to determine genes with significant differential expression in mutant strains and genes of interest for follow-up studies.

2.7.1 Differential expression analysis

The R package MicroArray Analysis of Variance model (**MAANOVA**) was used to test the probes for differential expression. The F_s statistic, which makes no assumptions about the distribution of variance but borrows information across the probes tested to determine the variance, was used as the test statistic, with a nominal *P*-value threshold of 0.05. The false discovery rate (FDR) for each probe was also determined with the **MAANOVA** *adjPval* function. Probes with a fold change of 1.3 (+/-) $p<0.05$ compared to the appropriate reference set were classed as being differentially expressed.

2.7.2 Functional Analysis (Including IPA Pathway analysis)

GO terms for each gene were collected from Mouse Genome Informatics (http://www.informatics.jax.org/) and Ensembl Genome Browser (http://www.ensembl.org/index.html) where available.

Pathway analysis was carried out using the Ingenuity Systems pathway analysis programme (http://www.ingenuity.com/) held on licence at the University of Edinburgh. All probes expressed above background for each strain were initially loaded into the system regardless of fold change or p-value. These two factors were then noted as observations and appropriate cut-offs (as mentioned in section 2.8.1) set before running the analysis. This resulted in the previously determined differentially expressed genes being highlighted and those expressed above background shaded grey indicating there presence but not significance. Any genes not expressed above background that appeared within the networks were white.

Ingenuity® works by accessing a database of gene interactions and comparing the input gene list with what is known by the database. A significance score is assigned for each

network and displayed as the negative log of the networks p-value. This indicates the likelihood that the assembly of focus genes within a network could be explained by random chance. A score of 3 indicates a 1 in 1000 chance that the focus genes are together by random chance so scores greater than this have at least 99.9% confidence that the genes are not grouped by random chance alone. Pathway analysis was carried out for both adult mouse mutants, and embryonic stages.

2.7.3 GOTree Analysis of differentially expressed genes

GOTreeMachine (GOTM) (http://bioinfo.vanderbilt.edu/gotm) is an online open resource for identifying clusters of genes by ontology. By comparing a list of genes of interest with a reference gene list it is possible to identify statistically enriched gene ontology categories by hypergeometric testing. This gives a list of genes which may have functional significance within the dataset and as a result within the disease group. I entered the complete list of differentially expressed genes from the microarray study as one data set and split by genotype and drug treatment to test if there was an overall over enrichment of specific gene sets within the list as a whole or within one group in particular.

2.7.4 Testing for overlaps and gene expression corrections from drug treatments

The resulting differentially expressed genes for each strain were compared using the *countif* function in Microsoft Office Excel. The adults were compared to each other and to their drug treated counterparts, and the embryonic stages were compared to each other resulting in 5 comparisons in total. Any overlaps in expression changes were noted as these may signify genes of interest for follow-up studies.

2.7.5 Comparing data with previously published work

The differential expression data was compared with previously published works on schizophrenia genetics to determine if any overlaps were present within our data. Gene lists from Walsh et al 2008 [83], Camargo et al 2007 [132] and Kirov et al 2008 [217] which have been implicated in schizophrenia were compared to my own lists generated by the differential expression analysis. Again any genes of overlap were noted for possible follow-up.

2.8 Follow up analysis of the microarray study

On completion of the microarray study it was necessary to validate the gene list using real time PCR. This was carried out over two stages to give a set of robust genes which would then be tested further using protein analysis.

2.8.1 Identifying the genes to take forward for validation

As previously mentioned the genes which were shown to be differentially expressed by the statistical analysis of the microarray data were then subject to functional analysis through IPA and GOTree, and were tested for overlap with previous key genetic studies of major mental illness. Literary searches were also completed to identify previous publications concerning these genes and whether they were significant to the study of major mental illness. Each gene was given a score based on the evidence available from the above analyses and its fold change and p value as noted from the ANOVA of the microarray. For each group the top 10-15% of genes after scoring were chosen for follow up by real time PCR.

2.8.2 Preparation of Samples for two phase follow up

As the follow up study was two phase there were two sets of samples to be prepared. The first phase involved the samples which had been used on the array. RNA from these mice was already stored at -80°C and cDNA synthesis carried out as outlined in section 2.2.6.

The second phase was to be carried out on an independent sample set of mice bred at the Biomedical Research Facility at the Western General Hospital, Edinburgh. These second phase mice were descendents of animals that had been shipped over from the Samuel Lunenfield Institute after the collection of the first phase samples. This second phase collection allowed me to use wild-type littermates, which had not been possible for the first round, and determine how useful the C57BL6/J was as a wild-type control in the initial study. Samples were processed in the same way as the first phase samples with the exception of the RNA extraction method. The PARIS procedure (section 2.2.2) was carried out on phase two mice to allow later protein analysis to be carried out on samples from the same mice used for the DNA study.

2.8.3 Identification of suitable probes for Taqman real time PCR

Real time PCR was to be carried out using the ABI taqman bioanalyser at the Wellcome Trust Clinical Research Facility. Once genes for follow up had been identified probes were ordered from Applied Biosystems. When choosing probes it was important to take into account the possible isoforms of any one gene and map the Taqman probe as closely to the original array probe as possible. The microarray probe sequences were provided with the output data from the CRF and I used these to map to the area of the gene targeted by the microarray. I then compared this to the area targeted by the various Taqman probes and picked the probe with the best overlap to ensure as much as possible the same isoforms of the gene were being detected. It was impossible to have direct overlap of both probes as Applied biosystems do not make the probe sequences available but identify gene regions targeted. I also tried to have probes with a _m1 product code

ending as this signified a probe which spans an exon junction and so would be more reliable with cDNA and cut out background from any possible genomic contamination.

Two control genes were required to allow for accurate comparisons to be made. The most commonly used control genes for mouse Taqman studies are *GAP-DH* and *Actin*. As *DISC1* is believed to play a role in the Actin cytoskeleton this was considered inappropriate as a control gene. *Hprt1* was chosen instead. This had previously been used successfully by the Cystic Fibrosis group within the centre and was known to be expressed in brain tissues.

2.8.4 Verification of genes by ABI Taqman qRT-PCR

Samples were analysed using the ABI Taqman machine at the Wellcome Trust Clinical Research Facility, Western General Hospital, Edinburgh. Samples were transported on ice and Taqman master mix was kept at 4°C within the CRF facility for my use.

The analysis was run on 396 well clear plates with two control genes and x genes of interest (x determined by how many samples were being run that day). A standard curve from pooled cDNA was run for each gene on each plate as an indicator of how well the reaction was working. A 1:200 calibrator was also run on each plate, allowing comparison between plates and within plates as this was used in the statistical analysis as a correctional value. Each well contained 5µl Taqman mastermix, 0.5µl appropriate probe and 4.5µl cDNA. This was sealed with a plastic sheet using a heat sealer and centrifuged for 30 seconds before being placed on the ABI Taqman platform. After completing the plate layout on screen the platform was set to run standard real time PCR cycle lasting 1.5 hours. On completion of the cycle the samples were removed from the platform and discarded, and the data analysed.

2.8.5 Statistical Analysis of ABI Taqman output

Samples were initially analysed in the ABI taqman screen to determine the efficiency of the reaction. For the replicated samples to be efficient the mean CT value for each replicate should be within 1cycle. Data was scanned to determine if this was the case and any samples which did not fall within these criteria were omitted. Results were then exported as a text file to be opened in excel for further analysis.

In excel the exported data was summarised to quantity mean and standard deviation of each sample. The normalisation factor was calculated by taking the geometric mean of the two control genes per sample and dividing by the average of the geometric mean, giving a unique normalisation factor for each sample. This was also done for the standard deviation using the calculation:

NF(sd) = NF*((control1 sd/2*(control1 average)) ^2 + (control2 sd/2*(control2 average) ^2) ^0.5

For each gene of interest this normalisation factor was used to create a normalised quantity mean for each sample by dividing the sample quantity mean by the normalisation factor. The normalised standard deviation was then calculated using the equation:

Norm sd = Normalised quantity mean*((NFsd/NF) ^2 + (Gene of Interest sd/Gene of Interest quantity mean^2) ^0.5

This normalised quantity mean and standard deviation was then rescaled to correct for the 1:200 calibrator by dividing the normalised quantity mean of the sample by the normalised quantity mean of the calibrator, and similarly with the standard deviations. It was these ‘rescaled’ values that would be used for further analysis. Outliers were detected using the inter-quartile range detection method and discarded from further analysis.

The group average, standard deviation and number were calculated using excel basic statistics and these values were transferred into a GraphPad Prism file for further analysis (www.graphpad.com). GraphPad Prism is a statistical analysis and graphing programme

on licence within the University. 'Raw' rescaled values were also transferred and used to create a scatterplot and carry out statistical analysis. For each gene of interest comparisons were made between appropriate groups using both unpaired t-tests and Mann-Whitney U tests. Which test was most appropriate for which gene was determined by drawing normal distribution curves for each gene of interest. Genes with normal distribution in samples were analysed by t-test, and non-normal distribution by Mann-Whitney U test. However, in most cases both tests gave similar results so the need to choose between them was arbitrary. Also with such relatively small sample sizes it is almost impossible to gain an accurate representation of normality distribution and this was used as more of a guide than a definitive distribution curve. Samples which maintained a p-value <0.05 were carried forward for further analysis.

Analysis of the *Disc1* developmental profile was carried out in the same way but graphed using rescaled averages with standard error rather than scatterplots. Unpaired t-tests and Mann-Whitney U tests were still carried out to determine the difference in *Disc1* expression at key time points between strains.

2.9 Protein analysis and Antibody Staining

Once the microarray validation was complete it was considered logical to look at protein expression both from protein lysates and cultured primary neurons.

2.9.1 Identification of suitable antibodies

Antibodies were ordered from Abcam UK and R and D systems, UK. Suitability for use was determined by searching the supplier website for antibodies which were known to work in western blotting and immunocytochemistry (ICC), and that had been shown to work in mouse samples.

2.9.2 Western Blotting

Composition of all buffers can be found in section 2.10.

Sample lysate (20µg in 10µl) was added to 10µl protein sample buffer and 1µl DTT and boiled for 5 minutes to denature. Due to the small product sizes all westerns were run on 7% polyacrilamide gels. Samples and marker were added to the gel in the presence of running buffer and a positive control of SHSY5Y cell lysate which was used in the initial antibody work up. The block was topped up with running buffer and set at a constant voltage of 150V for 90 minutes. Sponges were soaked in transfer buffer along with PVDF membrane which had previously been dipped in 100% methanol. Once the acrilamide gel had run for 90 minutes it was removed from the block and the case cracked open to allow access to the gel. The PVDF membrane was applied to the gel and filter paper placed either side. A transfer casket was loaded with 2 soaked sponges, then the filter paper – gel – membrane – filter paper sandwich, and finally another 2 sponges. The casket was loaded into the block and filled with transfer buffer while the surround was filled with 650ml cold dH_2O to prevent overheating during transfer. Blotter was set on constant 30V for 60 minutes. The membrane was removed and washed twice with dH_2O and then in Ponceau stain to visualise protein. The membrane was washed again with dH_2O and put on a shaking plate at low rpm until protein was visible before transferring to 40ml blocking buffer for 1hour on a shaking plate.

Primary antibodies were diluted in 4ml blocking buffer and applied to the membrane. Antibodies were diluted as per manufacturers instructions and a full list of conditions can be found in Appendix 1. Membranes were incubated at room temperature for 1 hour, or at 4°C overnight, on a low rpm shaking plate. Membranes were washed in wash buffer as previously described and the appropriate secondary antibody added to the membrane, diluted in 4ml wash buffer, for 25 minutes on a shaking plate. Membranes were washed in wash buffer as previously described and dabbed dry with tissue paper. ECL PLUS mixture (4.8ml reagent A + 120µl reagent B) was applied to the membrane and left to incubate for 5 minutes at room temperature. Membranes were dabbed dry and wrapped in clingfilm, ensuring there were no air bubbles in the film, before being placed in a cassette

to prevent bleaching. Light sensitive film was placed over the membrane in the cassette and sealed for the appropriate exposure time before being developed in a BioRad developer.

Films were kept for further analysis and the membranes stripped by placing in 10ml stripping buffer for 20 minutes on a shaking plate at room temperature. Membranes were then re-blocked and stained for GAP-DH using the same protocol as above.

2.9.3 Quantification of Westerns using Image J

Films were scanned into a PC in high resolution and imported into Image J for analysis. Image J reads the signal intensity of the bands and background within a chosen frame and returns a numerical output for each band. The signal intensities were corrected for loading differences by dividing the numerical output from the gene of interest by the signal intensity of GAPDH on the same membrane. The mean signal intensity and standard error for each group was calculated in Microsoft Excel and imported into GraphPad prism for further statistical analysis and graph production.

2.9.4 Culture of Primary Neurons

Coverslips were prepared by incubating overnight in 100% ethanol before washing with distilled water, heat-dry sterilisation and incubating in Poly-D-Lysine solution.

Adult female mice were sacrificed 18 days post conception and embryos removed for neuron culture. The E18 embryo brains were removed and placed in dissection buffer containing HBSS (+$CaCl_2$ +$MgCl_2$), L-glutamate and HEPES (1M, pH 7.3-7.5) and kept on ice through the dissection process. The hippocampus was removed and placed in fresh dissecting buffer under a standard dissection microscope. Once all hippocampi had been collected, the dissection buffer was removed and the tissue chopped up using a scalpel. Trypsin was added at 0.1% to aid protein digestion and incubated at 37°C for 45mins. The suspension was passed through a wide Pasteur pipette 5x to remove clumps and

centrifuged at 1500rpm for 5mins. Trypsin was removed and the pellet resuspended in 10ml DMEM with foetal bovine serum to deactivate the trypsin. The process was repeated to ensure all traces of trypsin were removed and then the suspension was passed through a wide necked Pasteur pipette 20x and a narrow necked Pasteur pipette 10x. Suspension was centrifuged at 1500rpm for 5 mins, supernatant removed and pellet resuspended in 10mls DMEM (no FBS). The suspension was passed through a 40µM cell filter and cells counted before being centrifuged and resuspended in Neurobasal +++ medium (Neurobasal medium, B-27 supplement, GlutaMAX-1 supplement, Pen/Strep solution) at a concentration of $2x10^5$ in a 12 well dish previously prepared with Poly-D-Lysine coverslips.

Cells were incubated at 37°C for 21days before being used for ICC. The neurobasal +++ medium was changed every 7days and cells checked to ensure health and growth.

2.9.5 ICC of Primary Neurons

The neurobasal medium was removed from the plate wells with an aspirator and coverslips washed twice with PBS at 4°C. Ice cold methanol was added to the coverslips and left to incubate at room temperature for 10minutes to fix the cells. Methanol was removed by aspiration and coverslips washed a further two times in PBS. Cells were blocked in 3% BSA for 20 minutes on a shaking plate at low rpm. Cells were then given two quick washes in PBS followed by 3 5 minute washes in PBS to remove all of the BSA. Primary antibodies were diluted in 3% BSA to a total volume of 200µl. After numerous trials a dilution of 1:50 was deemed the most appropriate for all primary antibodies of the genes of interest. Cells were also co-stained for alpha-tubilin and PSD-95. The antibodies were added to the coverslips and left on a shaking plate at room temperature for 1 hour to incubate. The primary antibody was removed by aspiration and coverslips washed as after the blocking treatment. Appropriate fluorescently labelled secondary antibodies were chosen and diluted in 3% BSA to a total volume of 200 µl. The colour of the fluorescent probe determined the dilutions. Red was diluted 1:800, green 1:500 and blue 1:1000. Secondary antibodies were added to the coverslips and left

to incubate on a shaking plate at room temperature for 1 hour. To prevent bleaching of the secondary antibodies the plate was wrapped in tinfoil to protect from light. The coverslips were washed as after the primary antibody treatment. A drop of mounting medium containing a Dapi stain was placed onto fresh slides and the coverslips mounted onto these with the cells facing downwards. Air bubbles were removed and slides stored away from light at 4°C until viewed with the confocal microscope.

2.10 Buffers

2.10.1 Western Blotting

Protein Sample Buffer (PSB)

- 6.25ml 1M Tris pH 6.8
- 10ml glycerol
- 10ml 20% SDS
- 13.75ml dH_2O
- 0.5mg Bromophenol Blue

Ponceau Stain

- 0.5g ponceau S
- 2ml Acetic acid
- 98ml dH_2O

Wash Buffer

- 100ml 10x PBS
- 900ml dH_2O
- 2ml tween

Running Buffer

- 20ml 20x NuPage Tris Acetate SDS running buffer

- 380ml dH_2O

Transfer Buffer

- 25nl NuPage transfer buffer
- 50ml methanol
- 425ml dH_2O

Blocking Buffer

- 2.5g Marvel
- 50ml PBS
- 100µl tween

2.10.2 Primary Neuron Production

500 µg/ml Stock Poly-D-Lysine Solution

- 5mg Poly-D-Lysine
- 10ml distilled water
- Filter sterilised and stored at -20°C

20 µg/ml Poly-D-Lysine Solution

- 25ml Borate buffer pH 8.5
- 22.5ml Distilled H_2O
- 2.5ml Stock Poly-D-Lysine

Dissection Buffer

- 500ml HBSS (+$CaCl_2$ + $MgCl_2$)
- 5ml L-glutamate
- 3.5ml HEPES (1M, pH7.3-7.5)

PBS

- PBS – (no $CaCl_2$ + $MgCl_2$)

0.1% Trypsin

- 20ml Trypsin/EDTA 0.25%
- 30ml dissection buffer

Neurobasal Medium

- 500ml Neurobasal medium
- 10ml B-27 supplement
- 5ml GlutaMAX-1 supplement
- 2.5ml Pen/Strep solution

DMEM

- 500ml DMEM
- 50ml foetal bovine serum
- 2.5ml Pen/Strep solution

Chapter 3
Sample collection and processing of a genome wide microarray analysis of the Disc1 ENU mouse mutants

3. Sample collection and processing of a genome wide microarray analysis of the Disc1 ENU mouse mutants

3.1 Introduction

DNA microarrays are widely considered the gold standard in genome wide expression studies of psychiatric genomics [99]. The complex multistep process allows extensive analysis of genome wide gene expression levels from relatively small sample numbers. This chapter describes the experimental preparation and preliminary analysis of a genome-wide microarray study comparing the Disc1 ENU mutant mouse lines described by Clapcote *et al* (2007)[197], with C57BL/6J control animals.

3.1.1 Background and motivation

Recently studies of ENU mutant mice have shown deficits in prepulse inhibition connected to genetic mutations, which can be partially reversed by antipsychotic medications[197]. Clapcote *et al* (2007) identified two missense mutant mouse lines with mutations within exon 2 of the Disc1 gene, in which they showed PPI deficits that were partially reversible with antipsychotic and antidepressant medication. These two mouse strains, L100P and Q31L, were described as 'schizophrenic-like' and 'depressive-like' by virtue of their behavioural profile and physiological observations and pharmacological responses that were considered comparable to human disease [218] (Table 3.1).

The two exon-2 mutants display two slightly different phenotypes corresponding to a schizophrenic-like and a depressive-like phenotype. The more severe schizophrenic-like phenotype occurs in the L100P mouse mutant. Here a point mutation has altered amino acid 100 from a leucine to a proline residue in the mouse Disc1. These mice have a marked reduction in PPI and overall startle response. The pronounced PPI deficit is partially rescued by administration of the typical antipsychotic haloperidol and the atypical antipsychotic clozapine. Interestingly, the anti-depressant and PDE4B inhibitor, rolipram also had a significant effect in rescuing PPI in the L100P mutant mice. Disc1 is known to bind to particular PDE4 isoforms at specific binding sites in the protein's head

domain[104]. Two of these binding sites are disrupted by the mutations in both mouse lines suggesting that binding to PDE4B is likely to be impaired. The Q31L mutant mouse also displayed a reduction in PPI, although not to the extent of the L100P mutant. Interestingly, the deficit displayed by the Q31L "depressive-like" mice was not reversible by treatment with antipsychotic medications or with rolipram. Consistent with the lack of response to rolipram treatment the Q31L mice also exhibit decreased PDE4B activity compared to wild-type littermates and L100P "schizophrenic-like" mutant mice [197]. The antidepressant bupropion abolished the PPI deficit in these animals.

Both mouse lines displayed a disruption in Latent Inhibition (LI), which is a phenomenon by which prior exposure to a stimulus that holds no reward decreases the response to that stimulus when it is paired with an unconditioned stimulus. Again, administration of clozapine abolished this effect in the L100P mouse mutant, but not in the Q31L line. The Q31L mouse mutant, but not the L100P mouse mutant, displayed increased immobility during a forced swim test, a measure of behavioural despair, and reduced social interaction and reward responsiveness (measures of withdrawl and anhedonia respectively). The Q31L and L100P mouse models had overall reductions in brain size of 6 and 13% respectively, with tissue contraction mainly of the cortex, entorhinal cortex, thalamus and cerebellum.

The aim of the work described in this and the following chapter was to partially replicate the study by Clapcote and Roder to confirm the presence of PPI deficits in the Q31L and L100P mouse lines when compared to my control group of C57BL/6J mice bought in from Jackson laboratories that were to be used in my microarray study, and to identify genes whose expression was altered as a result of these mutations. In the original behavioural study, administration of selected antipsychotic and antidepressant drugs partially rescued the behavioural phenotype of the two lines.

In this chapter I will describe an attempt to replicate the behavioural study and describe the sample collection and primary analysis of a genome wide gene expression study in the ENU mutant mouse lines. By identifying other genes dysregulated by these mutations

it may be possible to identify pathways involved in major mental illness and possible new targets for treatment. Millar et al previously described a dynamic interaction between DISC1 and PDE4B that served to regulate cAMP signalling (Millar 2005)[131]. Disruption of this interaction is predicted to alter modulation of cAMP signalling and may result in abnormal transcription of genes with cAMP response elements in regulatory regions. DISC1 also interacts with the cAMP response element-dependent transcription factor ATF4, and possibly binds to chromatin remodelling factors such as SMARCE1 [132]. These interactions with transcription factors, together with the fact that DISC1 localises to the nucleus [172, 173] is consistent with a role for DISC1 in transcriptional regulation. Thus the view was that examination of the gene expression profile, by whole-genome microarray analysis, in the developing and adult brain of the two Disc1 mutant mouse lines was a key step towards understanding the factors underlying the behavioural and anatomical features of this mouse model.

3.2 Partial Replication of Clapcote *et al's* Behavioural Analysis of the Disc1 ENU mouse mutants

Prepulse inhibition (PPI) is a neurological phenomenon in which the reaction to a startle stimulus (pulse) is reduced when it is preceded by a weaker prestimulus (prepulse) within a short time frame of under 500ms (figure 3.1). The reduction in startle response reflects the ability of the nervous system to adapt temporarily to a strong stimulus when a warning stimulus is present. It is generally accepted that this reflects the ability to filter out unnecessary information and so is a good measure of sensorimotor gating, a neurological function known to be deficient in schizophrenia[219].

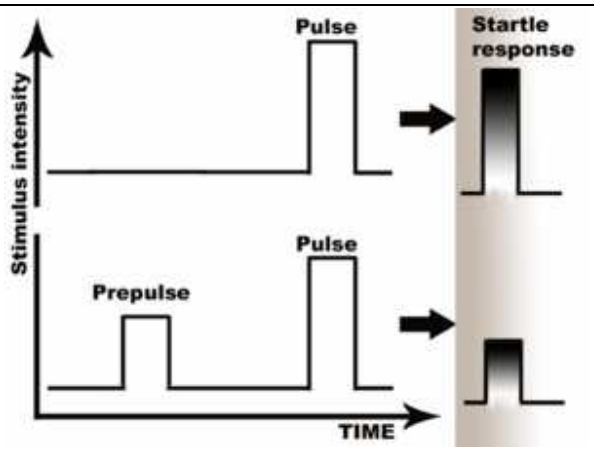

Figure 3.1: Image depicting the effective reduction in an animals startle response when a strong stimulus is preceded by a weaker prestimulus.

In some healthy individuals prepulse inhibition is as great as 70%. Individuals with major mental disorders such as Alzheimer's, schizophrenia, and bipolar depression however have significantly reduced prepulse inhibition compared to healthy controls[220]. As PPI deficits are observed in many mental disorders they are not diagnostic, but may indicate deficits in mental pathways common to these disorders. In particular, the inability to filter out unnecessary information could explain the symptoms of hallucinations and delusions associated with these disorders. In human patients there is also a noted gender difference in prepulse inhibition with males having higher PPI than females under all conditions[221]. Atypical antipsychotic medications such as risperidone and olanzapine are particularly effective in increasing PPI in individuals where a deficit is present[218].

Prepulse inhibition is not a purely human characteristic and has been widely observed in other mammals including rodents [219]. These animals are tested in a startle chamber (figure 2.4) with sensors detecting whole body movement as a measure of startle response. Pulse and pre-pulse are tones fed into the chamber via speakers mounted in the wall and controlled by computer programme. Pulse-alone results are compared to prepulse-plus-pulse, and the percentage of the reduction in the startle reflex represents prepulse inhibition. The baseline activity of each animal is also measured and subtracted

from the startle response measurements. Baseline activity does not appear to affect the prepulse inhibition of an individual[222].

	Change compared to wild type		
Behaviour	Q31L hom	L100P hom	
Anxiety	=	=	
Startle reactivity	=	reduced	
Spatial learning and memory	=	=	
Taste Sensitivity	=	n/a	
Olfactory Function	=	=	
Prepulse Inhibition	reduced	greatly reduced	SCHIZOPHRENIC - LIKE
Latent inhibition	greatly reduced	greatly reduced	
Social interaction	reduced	=	DEPRESSIVE - LIKE
Reward Responsiveness	reduced	=	
Immobility during forced swim test	increased	=	

Table 3.1: Behavioural Phenotypes of the L100P and Q31L Disc1 mutant mice. Table displays result from multiple behavioural tests carried out by Clapcote *et al* (2007) on the L100P and Q31L mutant mice. Reduction in social interaction, reward responsiveness and increased immobility during a forced swim test indicated a 'depressive-like'phenotype in the Q31L mutant mouse. Greatly reduced PPI and LI in the L100P mutant mouse indicate a 'schizophrenic-like' phenotype.

Groups of mutant and control mice were also treated with drugs that had previously been successful in rescuing PPI in these mouse lines.

3.2.1 Effect of Gender on PPI in non-drug treated animals

Mice were tested as outlined in section 2.4.1. Twelve mice per group (8 Q31L adult) were tested on consecutive days. Analysis was carried out on startle response (120dB) and prepulse inhibition at 3 decibel levels (69, 73 and 81dB) and combined. The effects of gender, drugs and genotype were investigated.

There was no gender effect on total PPI ($F=1.393$, $p=0.226$) or acoustic startle response at 120dB minus pulse stimulus ($F=2.357$, $p=0.069$) in any group. As no gender differences were observed, further analyses were carried out grouping animals purely by genotype and drug treatment.

3.2.2 Effect of Genotype on PrePulse Inhibition and Startle Response in Mice

I found no difference in total prepulse inhibition ($F=1.788$, $p=0.185$) or acoustic startle response at 120dB minus pulse stimulus ($F=2.091$, $p=0.142$) by genotype in non-drug treated animals (Figure 3.2). It should be noted at this point, however, that comparison with previously published data showed that the mutant mice were displaying comparable levels of PPI and startle response as previously described, but that the wild type controls were displaying lower levels than previously reported. When the total prepulse inhibition for the C57BL6/J mice used in my study was compared to that of the wild type littermate controls in the previously published study they were significantly different ($t=4.07$, $p=0.0003$, $df=31$). This is also true for the acoustic startle response ($t=2.63$, $p=0.014$, $df=26$). If I substitute the previously reported wild type littermate data for my C57BL6/J controls in my comparisons, I find no significant differences except in one comparison, total PPI in the Q31L mutant ($t=2.17$, $p=0.029$, $df=27$).

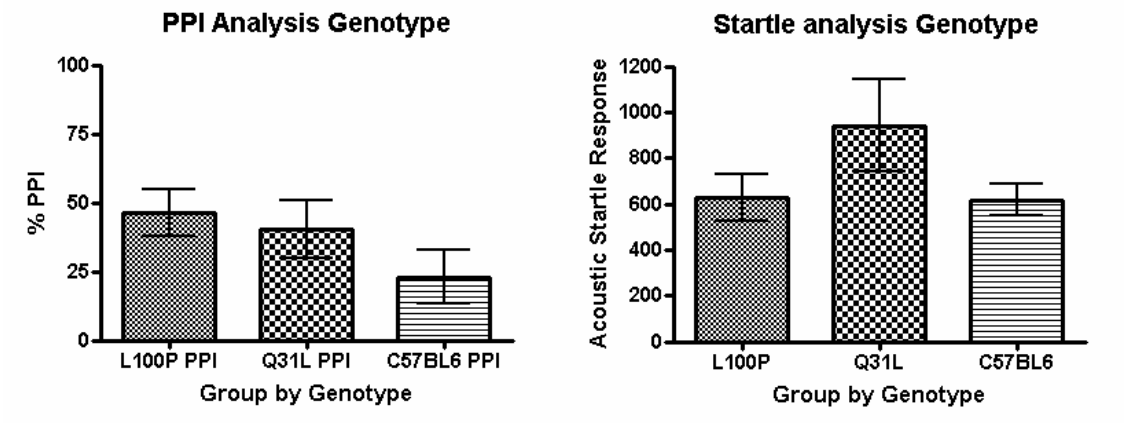

Figure 3.2: Total PrePulse Inhibition and Acoustic Startle Response in non-drug treated adult mice from each line. No difference is observed between groups in either comparison. Error bars depict standard error.

Clapcote *et al* had previously reported that the level of prepulse may be significant in the manifestation of PPI in the mutant mouse lines. Analysis of variance was carried out on the data from the individual prepulse conditions (69dB, 73dB and 81dB) to determine if this was the case with my mice. While the both the L100P and Q31L mouse lines displayed a trend with reduced PPI at 69dB which increased through 73dB and 81dB there was again no significant differences between groups (F=1.478, p=0.844).

	% PrePulse Inhibition non-drug treated mice		
	C57BL6	L100P	Q31L
69 dB	25.538 +/- 10.9 (n=12	35.597 +/- 9.9 (n=12)	38.711 +/- 10.4 (n=8)
73 dB	21.146 +/- 11.5 (n=12)	48.231 +/- 9.6 (n=12)	39.667 +/- 10.6 (n=8)
81 dB	22.322 +/- 9.3 (n=12)	55.235 +/- 6.45 (n=12)	42.681 +/- 11.6 (n=8)
TOTAL	23.002 +/- 9.9 (n=12)	46.354 +/- 8.3 (n=12)	40.353 +/- 10.5 (n=8)

Table 3.2: Mean results on PPI at all 3 levels tested, plus total. Pre-pulse inhibition was tested at 3 decibel levels, and averaged across all levels to produce a mean PPI total. Values shown are startle response and standard error.

3.2.3 Effect of Drug treatment on PrePulse Inhibition and Acoustic Startle Response in Mice

Prior to testing, 12 mice selected at random (8 Q31L) within each group had been treated with clozapine, bupropion or rolipram to determine the effects of these compounds on PPI and startle response within and between lines. Mice not receiving drug treatment received a saline injection.

	Acoustic Startle Response of Drug Treated Mice			Prepulse Inhibition of Drug Treated Mice		
	C57BL6	L100P	Q31L	C57BL6	L100P	Q31L
Saline	618.775 +/- 67.7 (n=12)	625.725 +/- 102.6 (n=12)	941.825 +/- 200.8 (n=8)	23.002 +/- 9.9 (n=12)	46.354 +/- 8.3 (n=12)	40.353 +/- 10.5 (n=8)
Clozapine	713.575 +/- 96.6 (n=12)	532.175 +/- 99.2 (n=12)	387.238 +/- 109.7 (n=8)	54.860 +/- 4.8 (n=12)	51.467 +/- 10.1 (n=12)	34.181 +/- 17.8 (n=8)
Rolipram	808.175 +/- 90.1 (n=12)	704.425 +/- 111.3 (n=12)	-	54.225 +/- 8.1 (n=12)	46.501 +/- 8.5 (n=12)	-
Bupropion	610.533 +/- 101.7 (n=12)	620.425 +/- 76.2 (n=12)	817.725 +/- 143.8 (n=8)	22.489 +/- 14.2 (n=12)	21.385 +/- 10.99 (n=12)	40.964 +/- 4.7 (n=8)

Mean acoustic startle response at 120dB stimulus, and average PPI over 69, 73 and 81 dB ranges

Table 3.3: Mean acoustic startle response and average PPI over all ranges tested. Table shows the total mean startle response and PPI by genotype and drug treatment. Values displayed are mean and standard error.

The effects of the drug treatment by genotype were analysed using a one-way ANOVA and post-hoc Tukey test. Total PPI was found to be significantly different in L100P drug treated mutants ($F = 4.67$, $p = 0.019$) and post-hoc analysis revealed clozapine treatment significantly decreased PPI, while bupropion treatment increased PPI in these mice. Acoustic startle response was not altered significantly in these comparisons. In the Q31L mutant no significant drug effect on PPI was found ($F=0.118$, $p=0.889$), however treatment with clozapine was found to significantly lower the acoustic startle response in these mice ($p=0.02$).

3.3 Whole Genome Gene Expression Study Experimental Design

3.3.1 Sample size and power

The RNA used in the microarray study was extracted from hippocampal brain tissues taken from the ENU mouse mutants immediately post behavioural testing. The result was 57 samples across six adult groups, as outlined in table 3.4.

Biological Group	Group Symbol	Number of Samples
C57BL/6J untreated	C57BL6	10
L100P untreated	L100Psal	9
Q31L untreated	Q31Lsal	10
L100P rolipram treated	L100Proli	8
L100P clozapine treated	L100Pcloz	9
Q31L bupropion treated	Q31Lbupro	11

Table 3.4 Summary of Adult Mouse Groups used in the genome-wide expression analysis. The groups are based on both genotype and presence or absence of drug treatment prior to behavioural testing. Group symbol is the annotation used throughout the rest of this thesis and the number of samples per group is shown in the final column.

Additionally, to test the effect of the Disc1 ENU mutations on genome-wide expression levels during development, 29 samples from three further groups were added from mouse Embryonic day 13.5 as outlined in table 3.5. Previously published data by Shurov *et al* (2004) [134] had shown Disc1 to have a peak in expression at this developmental time point, which coincides with neural determination and differentiation after a peak in neuronal migration at timepoint E12.5[223].

Biological Group	Group Symbol	Number of Samples
C57BL/6J Embryonic day 13.5	C57BL6 E13	12
L100P Embryonic day 13.5	L100P E13	7
Q31L Embryonic day 13.5	Q31L E13	10

Table 3.5 Summary of developmental mouse groups used in the genome wide expression analysis. Groups are based on genotype alone, no drug treatments were used in developmental samples. Group symbol is the annotation used throughout the rest of this thesis and the number of samples per group is shown in the final column.

The aim of this study was to determine which genes are dysregulated on the basis of genotype and drug treatment by comparing each of the ENU groups to their appropriate control groups. The C57BL/6J wildtype group was used as the control for testing the effect of genotype while the effect of drug treatment compared the drug treated group to their respective saline treated counterparts.

Power calculations were performed using the R/BioConductor (www.bioconductor.org) freeware with the *power.multi* function within the **sizepower** package[224]. As the true variability among individuals and actual effect size were unknown, a range of effect sizes and variance parameters were used to determine the ability of this experimental design to detect genes with differential expression. Natural variation among mice in inbred strains is thought to be very low (around 0.1-0.2 SD) so should result in low noise in the experiment and reasonable power to detect significant differential expression. As the mice used in this study had been backcrossed on to C57BL/6J (Jackson Laboratories) for 10 generations they could be considered an inbred strain and thus should have similar genetic variation. Standard deviations tested ranged from 0.1-0.5. It was, however, considered highly likely the natural variation between mice would be below the 0.2 SD range. No adjustments were made for multiple testing and five percent Type 1 error rate was assumed for each gene.

The power curves for the full range of SDs tested are shown in Figure 3.3. A sample size of 7 (as graphed) gives 80% power to detect a 1.3 fold change, the lower limit of the

illumina array platform, given an SD of 0.15. This sample size also provides 80% power to detect a two-fold change in more variable genes with SD of up to 0.4. This was the smallest group used, with most groups having sample sizes of 10-12 animals.

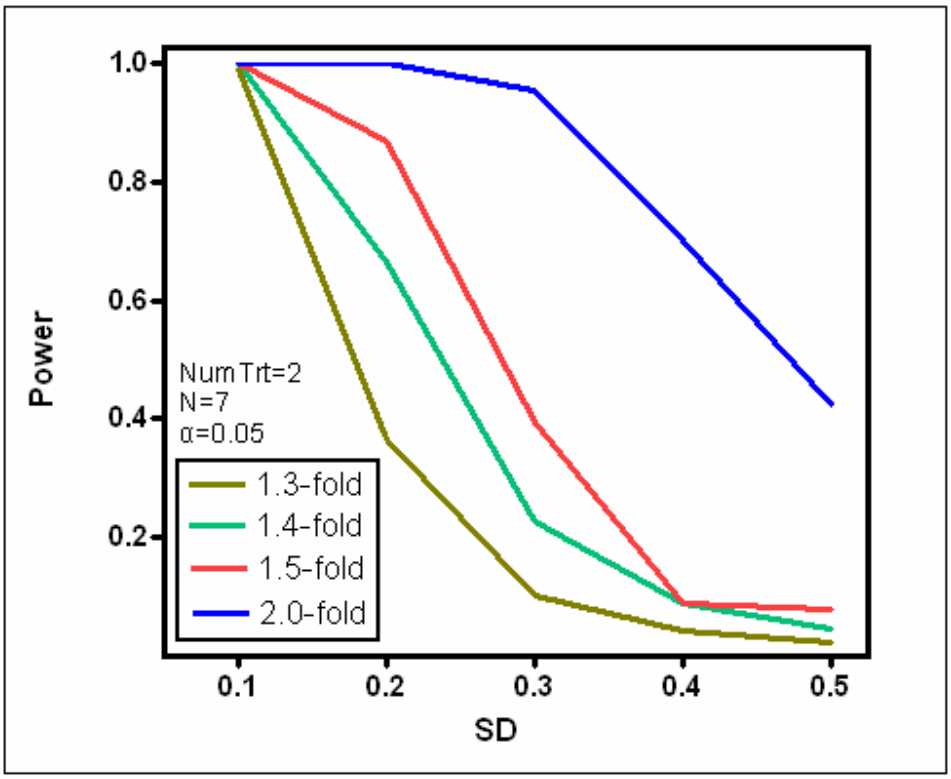

Figure 3.3: Power Curves. Power calculations were performed in the R/BioConductor (www.bioconductor.org) package **sizepower** using the *multi.power* function. Power curves were generated using the GraphPad software Prism 4.0 on license to the University of Edinburgh. The power (y-axis) was determined for the minimum number of samples and plotted against a range of standard deviation values (x-axis, range 0.1-0.5) using pairwise comparisons (NumTrt=2) with 7 samples per group (N=7, the smallest group size used) assuming a Type 1 error rate (α) of 0.05. Each coloured line represents the effect size expressed as fold change.

3.3.2 Sample Layout and Technical Replication

In total there were 86 samples to be hybridized to Illumina Mouse Ref-6 Beadchips. Due to the large number of samples available, and potential costs involved in running each sample individually, it was decided to pool samples based on genotype, drug treatment and gender to reduce the number of bead array chips required for the study. Pooling of samples resulted in 33 groups to be hybridized to the bead chip (figure 3.4) plus 9 replicate samples to be used for quality control. Pools from each biological group were evenly dispersed throughout the 7 BeadChips used to ensure that systematic bias or failure of one BeadChip did not sacrifice a whole biological group.

The nine replicate samples previously mentioned were used to assess the reliability between-BeadChip and within-BeadChip with different amplification and labelling preparations. This was an essential quality control method as comparisons would be made of data both between and within BeadChips. Firstly the non-drug treated C57BL/6J adult mouse female pool 2 (C57BL/6J-Sal F2) was hybridized to each BeadChip in the array in position A (figure 3.4) to allow assessment of the reproducibility between BeadChips. Secondly the same mouse RNA was amplified and labeled in three separate reactions (C57BL/6J-Sal F2 (1-3)) and hybridized to the same BeadChip in three different positions to assess the within-BeadChip experimental reproducibility. According to the BeadChip specifications (www.illumina.com), the expected array-to-array coefficient of variation (the ratio of the standard deviation of the mean expressed as a percentage and an indicator of within-BeadChip reliability) is <10%. This has been further corroborated by the Microarray Quality Control Consortium[225] through an intra-platform reproducibility study. As this has shown that within BeadChip reproducibility is reliable the main function of the technical triplicate was to determine the reproducibility and reliability of the amplification process and eliminate technical errors.

Array	Beadchip						
	1	2	3	4	5	6	7
A	***C57BL6 Sal F-2***	***C57BL6 Sal F-2***	***C57BL6 Sal F-2***	***C57BL6 Sal F-2***	***C57BL6 Sal F-2***	***C57BL6 Sal F-2***	*C57BL6 Sal F2(2)*
B	C57BL6 Sal F-1	L100P Cloz F-1	Q31L E13 F-2	C57BL6 Sal M-2	L100P Roli F-2	L100P Sal F-2	*C57BL6 Sal F2(3)*
C	L100P Cloz F-2	L100P Sal F-1	C57BL6 Sal M-1	L100P Cloz M-2	C57BL6 E13 F-2	C57BL6 E13 M-2	*C57BL6 Sal F2(1)*
D	L100P Roli M-2	Q31L Sal F-1	L100P Cloz M-1	L100P Sal M-2	Q31L Bupro F-2	L100P Sal M-1	L100P E13 M-1
E	Q31L Bupro M-1	Q31L E13 M-1	Q31L Bupro F-1	Q31L Sal F-2	C57BL6 E13 F-1	L100P Roli F-1	C57BL6 E13 M-1
F	Q31L E13 M-2	Q31L Sal M-2	L100P Roli M-1	L100P E13 F-1	L100P E13 F-2	Q31L E13 F-1	Q31L Sal M-1

Key			
C57BL6	Wild type non-littermate control		
L100P	ENU mouse mutant with L100P Disc-1 mutation 'schizophrenic-like' phenotype		
Q31L	ENU mouse mutant with Q31L Disc-1 mutation 'depressive-like' phenotype		
Sal	Non-drug treated	E13	Embryonic day 13
Cloz	Clozapine treated	F#	Female pool #
Roli	Rolipram Treated	M#	Male pool #
Bupro	Bupropion treated		

Figure 3.4: Layout of the seven Illumina mouse ref-6 version 2 BeadChips. Each column (1-7) represents a BeadChip and each row (A-F) represents an array on the BeadChip. Samples are named by their genotype, drug treatment, gender and pool number (as indicated in the key above). Samples in bold italics are the same sample on each BeadChip acting as between-BeadChip replicate controls. Samples in italics are from the same source RNA, amplified in three separate batches and hybridized to the array as within-BeadChip experimental reproducibility controls.

3.4 Sample Preparation and Quality

Mice from the previously described behavioural study were sacrificed immediately post PrePulse inhibition testing by cervical dislocation. The brain was removed and stored in RNA*later* for one week at 4°C to stabilize the tissues. Hippocampi were removed and RNA extracted. The time and date of tissue and RNA extraction was logged to allow identification of any batch effects. Total RNA was isolated and treated with DNase while on column to remove any contamination from genomic DNA.

3.4.1 Integrity of total RNA

Previous studies have shown that the integrity of total RNA is critical for successful microarray analysis. The Agilent 2100 Bioanalyser (Nano-LabChip) with 2100 expert software provides a standardized measure of RNA quality, the RNA Integrity Number (RIN) [226] which is based on the curve of the electrograph produced by the analyzer software package. RIN numbers range from 1 (totally degraded RNA) to 10 (totally intact RNA). A high RNA RIN number is obviously desirable however previous publications have advised that consistency of RIN values across samples is more important than high values[227]. A comparison between RNA integrity and success of GeneChip (Affymetrix) microarray analysis [228] suggests that RIN>5.5 gives sufficient quality of expression data. With both these points in mind, samples with RIN numbers below 5.5 were not discarded immediately but were retained to be used if the other samples within that group were of consistently low value. Table 3.6 lists the identification numbers of mice used in this study alongside the date of RNA preparation and the RIN number returned from the Agilent Bioanalyser. RIN values ranged from 5.2 to 10.0, with an overall median RIN of 9.2 and a Mode RIN of 10.0. Median RINs from the three embryonic groups were comparable (C57BL6 E13, 9.85; L100P E13 9.80; Q31L E13 9.55), within the L100P mutant group the rolipram treated pools stood out, with a lower RIN than the others (L100P Sal 9.05; L100P Cloz 9.55; L100P Roli 7.65) and within the Q31L mutant group the bupropion treated pools stood out, with a lower RIN than the others (A31L Sal 9.30; Q31L Bupro 6.50). The C57BL6 Sal pools had a comparable median RIN to the two mutant untreated groups (C57BL6 Sal 9.30). Representative electrophorographs are shown in figure 3.5.

(A) **(B)**

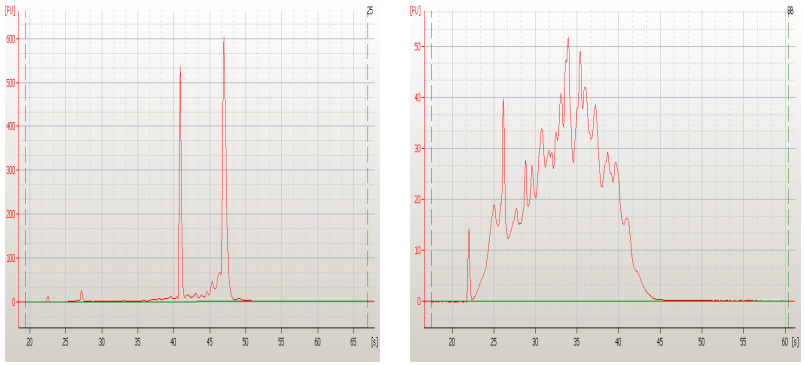

Figure 3.5: Examples of Agilent 2100 Bioanalyser output Electropherographs. (A) Total RNA with RIN 10 – intact with no contamination. (B) Total RNA with RIN 2.1 – low quality, high contamination and little intact RNA. Samples with RIN's this low would not be used.

Sample ID	Contained in Pool	Mutation	Drug treatment	Sex	RIN
C57-15-24	C57BL6 Sal F-1	Nil	SALINE	F	9.40
C57-14-24	C57BL6 Sal F-1	Nil	SALINE	F	9.20
C57-14-25	C57BL6 Sal F-1	Nil	SALINE	F	10.00
C57-13-24	C57BL6 Sal F-2	Nil	SALINE	F	8.90
C57-12-24	C57BL6 Sal F-2	Nil	SALINE	F	9.30
C5713-25	C57BL6 Sal F-2	Nil	SALINE	F	10.00
C57-9-4	C57BL6 Sal M-1	Nil	SALINE	M	10.00
C57-11-5	C57BL6 Sal M-1	Nil	SALINE	M	6.40
C57-8-4	C57BL6 Sal M-2	Nil	SALINE	M	6.3
C57-8-5	C57BL6 Sal M-2	Nil	SALINE	M	5.6
C57-10-5	C57BL6 Sal M-2	Nil	SALINE	M	9.60
31B-1-23	Q31L Sal F-1	Q31L	SALINE	F	9.50
31B-1-24	Q31L Sal F-1	Q31L	SALINE	F	10.00
31A-1-33	Q31L Sal F-2	Q31L	SALINE	F	8.90
31A-1-34	Q31L Sal F-2	Q31L	SALINE	F	9.80
31B-1-7	Q31L Sal M-1	Q31L	SALINE	M	9.20
31B-1-2	Q31L Sal M-1	Q31L	SALINE	M	9.00
31B-1-6	Q31L Sal M-2	Q31L	SALINE	M	8.80
31C-1-4	Q31L Sal M-2	Q31L	SALINE	M	9.40
31A-1-28	Q31L Bupro F-1	Q31L	BUPROPION	F	5.6
31B-1-25	Q31L Bupro F-1	Q31L	BUPROPION	F	5.2
31A-1-26	Q31L Bupro F-2	Q31L	BUPROPION	F	7.40
31A-1-29	Q31L Bupro F-2	Q31L	BUPROPION	F	6.1
31C-3-1	Q31L Bupro M-1	Q31L	BUPROPION	M	9.20
31A-1-9	Q31L Bupro M-1	Q31L	BUPROPION	M	6.6
31A-1-11	Q31L Bupro M-1	Q31L	BUPROPION	M	6.5
100C-1-21	L100P Sal F-1	L100P	SALINE	F	5.90
100B-1-31	L100P Sal F-1	L100P	SALINE	F	7.70
100A-1-28	L100P Sal F-1	L100P	SALINE	F	9.40
100D-1-23	L100P Sal F-2	L100P	SALINE	F	8.80
100C-1-22	L100P Sal F-2	L100P	SALINE	F	6.70
100B-1-32	L100P Sal F-2	L100P	SALINE	F	8.90
100B-1-3	L100P Sal M-1	L100P	SALINE	M	9.90
100D-1-4	L100P Sal M-1	L100P	SALINE	M	9.40
100A-2-2	L100P Sal M-1	L100P	SALINE	M	9.00
100B-1-7	L100P Sal M-2	L100P	SALINE	M	9.10
100A-1-3	L100P Sal M-2	L100P	SALINE	M	9.80
100D-1-3	L100P Sal M-2	L100P	SALINE	M	9.70
100B-1-27	L100P Cloz F-1	L100P	CLOZAPINE	F	9.40
100B-1-28	L100P Cloz F-1	L100P	CLOZAPINE	F	9.80
100A-1-27	L100P Cloz F-1	L100P	CLOZAPINE	F	9.80
100A-1-26	L100P Cloz F-2	L100P	CLOZAPINE	F	9.70
100D-1-21	L100P Cloz F-2	L100P	CLOZAPINE	F	9.20
100D-1-22	L100P Cloz F-2	L100P	CLOZAPINE	F	9.40
100B-1-1	L100P Cloz M-1	L100P	CLOZAPINE	M	10.00
100B-1-5	L100P Cloz M-1	L100P	CLOZAPINE	M	8.90
100D-1-1	L100P Cloz M-1	L100P	CLOZAPINE	M	10.00
100D-1-2	L100P Cloz M-2	L100P	CLOZAPINE	M	8.80
100B-1-6	L100P Cloz M-2	L100P	CLOZAPINE	M	8.10
100A-1-2	L100P Cloz M-2	L100P	CLOZAPINE	M	9.80
100C-2-24	L100P Roli F-1	L100P	ROLIPRAM	F	7.80
100D-1-26	L100P Roli F-1	L100P	ROLIPRAM	F	9.20
100A-2-27	L100P Roli F-2	L100P	ROLIPRAM	F	7.20
100C-2-23	L100P Roli F-2	L100P	ROLIPRAM	F	8.90
100B-4-10	L100P Roli M-1	L100P	ROLIPRAM	M	7.80
100C-1-2	L100P Roli M-1	L100P	ROLIPRAM	M	7.20
100B-4-5	L100P Roli M-2	L100P	ROLIPRAM	M	7.50
100C-1-1	L100P Roli M-2	L100P	ROLIPRAM	M	6.90
L100P-E13-2	L100P E13-M1	L100P	Nil	M	8.1
L100P-E13-3	L100P E13-F1	L100P	Nil	F	10
L100P-E13-5	L100P E13-M1	L100P	Nil	M	9.8
L100P-E13-7	L100P E13-F1	L100P	Nil	F	10
L100P-E13-8	L100P E13-F1	L100P	Nil	F	9.9
L100P-E13-10	L100P E13-F2	L100P	Nil	F	10
L100P-E13-11	L100P E13-F2	L100P	Nil	F	9.2
L100P-E13-12	L100P E13-M1	L100P	Nil	M	9.8
L100P-E13-13	L100P E13-F2	L100P	Nil	F	10
L100P-E13-15	L100P E13-M2	L100P	Nil	F	8.3
Q31L-E13-1	Q31L E13-M1	Q31L	Nil	M	8.7
Q31L-E13-2	Q31L E13-M1	Q31L	Nil	M	10
Q31L-E13-3	Q31L E13-M1	Q31L	Nil	M	9.9
Q31L-E13-5	Q31L E13-M2	Q31L	Nil	M	7.6
Q31L-E13-6	Q31L E13-F1	Q31L	Nil	F	9.5
Q31L-E13-7	Q31L E13-M2	Q31L	Nil	M	8.2
Q31L-E13-9	Q31L E13-M2	Q31L	Nil	M	9.7
Q31L-E13-10	Q31L E13-F1	Q31L	Nil	F	9.6
Q31L-E13-15	Q31L E13-F2	Q31L	Nil	F	9.9
Q31L-E13-16	Q31L E13-F2	Q31L	Nil	F	9.1
C57Crl E13-3	C57BL6 E13-M1	Nil	Nil	M	9.1
C57Crl E13-8	C57BL6 E13-F1	Nil	HOM	F	9.7
C57Crl E13-9	C57BL6 E13-F1	Nil	HOM	F	9.8
C57Crl E13-10	C57BL6 E13-F2	Nil	HOM	F	10
C57Crl E13-12	C57BL6 E13-F2	Nil	HOM	F	9.8
C57Crl E13-13	C57BL6 E13-M2	Nil	HOM	M	9.7
C57Crl E13-14	C57BL6 E13-M1	Nil	HOM	M	10
C57Crl E13-15	C57BL6 E13-M2	Nil	HOM	M	9.9

Table 3.6: Mouse samples used in the microarray analysis. Column 1 gives the animal ID, column2 the genotype and drug treatment, plus pool the animal was assigned to (F1 = female pool 1 etc), column 3 the mutation, column 4 the drug treatment, column 5 the gender and column 6 the RNA RIN number of the sample as an assessment of quality.

3.4.2 Integrity of cRNA

The RNA from the mouse samples in table 3.6 was used to produce biotinylated amplified cRNA (antisense mRNA). Labelled samples were produced within the two weeks prior to the microarray being run so no samples were stored over long periods. Quality and concentration of the samples were obtained using the Agilent Bioanalyser (Nano-LabChip). The Agilent Bioanalyser software does not provide a metric measure of quality for mRNA but rather reports on contamination levels by ribosomal RNA (rRNA) and provides an electropherograph of peaks in RNA size distribution that must be examined manually (figure 3.6). Evidence of a shift on RNA size distribution towards lower sized fragments would indicate degradation of RNA, which was not seen in any of the electropherographs analysed (figure 3.6). There were, however, some cases where contamination with rRNA was reported by the Agilent Bioanalyser which was not evident on manual examination of the traces. After communication with Agilent technical support, it was agreed the contamination reported was an artifact of the Agilent 2100 Expert software caused by high concentration levels close to the limits of detection for the Agilent Bioanalyser (Agilent RNA 6000 Nano chip = 25-250ng/µl mRNA).

Dilution of highly concentrated samples and re-quantification gave concentration measurements with a CV of less than 10% compared to the original undiluted measurement (Christoforou, personal communication). The Agilent Bioanalyser product specifications expect a CV of less than 10% for any reproduced quantification measurement making this well within the expected reproducibility of sample quantification. For this reason it was decided not to re-quantify these samples and to use the original quantification results provided the electropherograph traces showed no evidence on contamination peaks at 18S and 28S (the fragments which make up total RNA). This allowed me to be confident that the samples submitted to the Wellcome Trust Clinical Research Facility, Western General Hospital, Edinburgh, for hybridization to the array chips were of sufficient quality and accurately quantified.

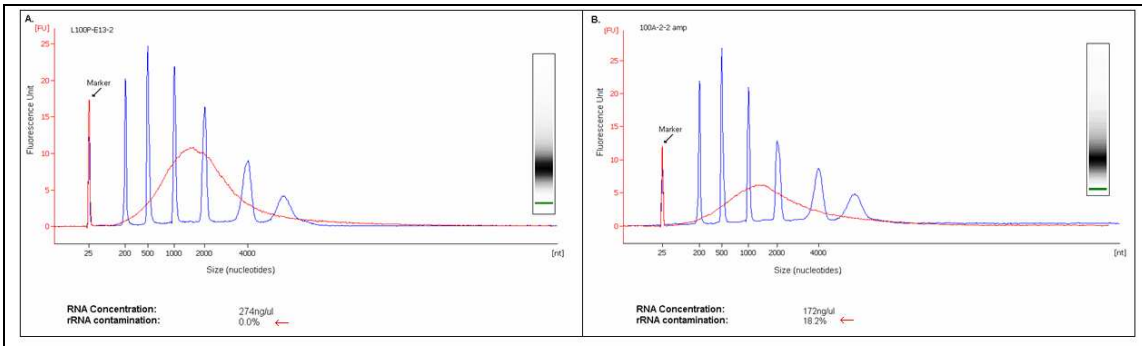

Figure 3.6: Agilent Bioanalyser output of cRNA Integrity. Figure shows examples of two samples that were analysed using the Agilent Bioanalyser 2100 Expert software. Fluorescent units (FU) are plotted against size of RNA fragment (nucleotides). The RNA marker is indicated in the trace and on the gel image on the right hand side of each trace. RNA concentration and contamination are indicated under each graph. Trace A shows a sample with good concentration and no estimated contamination while trace B shows good concentration but predicted 18.2% contamination. As seen on the trace there is no evidence of a peak at 18S or 28S suggesting this is an artifact rather than true contamination.

3.5 Data Generation using the Illumina BeadArray Platform

Hybridisation of the cRNA samples was carried out at the Wellcome Trust Clinical Research Facility (WTCRF) at the Western General Hospital in Edinburgh. Samples were hybridized to seven Illumina Mouse-6 version 2 BeadChips on the Illumina BeadArray Platform. The chips were scanned using the Illumina BeadArray reader software, and image processing steps to determine the bead signal data from the raw images was automatically performed [229]. The Illumina BeadArray platform uses multiple bead copies for each probe allowing each probe signal intensity to be estimated from a number of beads reducing the likelihood of complete probe failure. Each bead copy is randomly positioned within the Beadchip to negate regional bias, and on average each probe signal intensity is estimated from readings from 30-40 bead copies with a low probability of being calculated from less than five beads [229]. Outlying beads (those with a median deviation more than 3) were removed and the average intensity of the remaining beads used to calculate the probe signal. Across the seven chips used in this experiment the

average number of beads per probe was 37.8 with one probe intensity being calculated from less than 5 beads (table 3.7). Illumina BeadStudio software was used to generate the probe signal intensity from multiple bead readings, along with the standard deviation of bead intensities. The detection p-value (which determines whether a probe is present in the given sample) was also calculated using this software.

Of the 48,000 probes on the BeadChip, 46,644 probes were successful (corresponding to 34,492 individual genes). For the purposes of quality control, preprocessing and differential expression analysis, each probe was identified not only by probe ID but by gene name, meaning some genes appear more than once in the initial stages or analysis. This is purely because some genes had multiple probes and as such could not be identified by gene name alone.

Table 3.6 shows the estimated average intensity and intensity range for each array as well as the average number of beads and minimum number of beads used to calculate the probe intensities. Across the whole experiment the mean probe intensity was 551.7 with a range of intensities from 10.1 to 48574. Interquartile range analysis was used to confirm no BeadChip probe intensity was considered an outlier and a one-way ANOVA (with 41 degrees-of-freedom) was performed on the raw, $\log_2$-transformed data to account for variability in the data. As expected the ANOVA showed statistically significant variation among the 42 arrays ($F=5.23$, $p=0.0006$), due to the effect of non-biological factors such as RNA integrity, hybridization and scanning, validating the need to normalize the data prior to further analysis.

Sample ID (BeadChip)	Average # Beads (min)	Average signal (range)	Background (noise)
C57BL6Sal F2 (1)	37.8 (9)	548 (35.7-45324)	67.9 (20.3)
C57BL6Sal F1 (1)	38.0 (10)	531 (31.1-40870)	66.3 (19.6)
L100PCloz F2 (1)	38.1 (6)	634 (36.6-45103)	73.0 (22.0)
L100PRoli M2 (1)	37.3 (5)	604 (36.4-44599)	74.5 (22.5)
Q31LBupro M1 (1)	37.4 (6)	540 (37.1-41067)	68.7 (19.8)
Q31LE13 M2 (1)	37.9 (9)	606 (37.5-43062)	71.6 (24.1)
C57BL6Sal F2 (2)	37.9 (11)	615 (42.4-48574)	76.0 (21.2)
L100PCloz F1 (2)	37.6 (9)	727 (35.3-45952)	81.9 (26.0)
L100PSal F1 (2)	36.4 (7)	487 (34.9-44619)	67.3 (18.2)
Q31LSal F1 (2)	38.2 (8)	575 (32.3-45039)	69.5 (19.9)
Q31L M1 (2)	38.01 (7)	601 (35.6-45445)	71.8 (20.6)
Q31LSal M2 (2)	37.5 (9)	603 (36.5-43507)	71.5 (21.8)
C57BL6Sal F2 (3)	37.8 (9)	607 (37.0-41721)	75.2 (22.6)
Q31LE13 F2 (3)	37.7 (9)	563 (36.3-42270)	68.9 (17.7)
C57BL6Sal M1 (3)	38.1 (10)	543 (34.4-46216)	67.8 (18.5)
L100PCloz M1 (3)	38.1 (8)	537 (35.5-40947)	67.7 (19.3)
Q31LBupro F1(3)	38.1 (8)	467 (34.9-39263)	63.7 (17.7)
L100PRoli M1 (3)	37.5 (9)	534 (33.2-42955)	67.2 (19.1)
C57BL6Sal F2 (4)	37.7 (7)	574 (34.8-48172)	68.6 (19.6)
C57BL6Sal (4)	37.9 (8)	581 (26.3-41383)	66.0 (18.3)
L100PCloz M2 (4)	37.8 (8)	506 (33.3-45198)	61.8 (15.6)
L100PSal M2 (4)	38.2 (9)	622 (32.6-42080)	68.1 (19.5)
Q31LSal F2 (4)	38.1 (7)	571 (31.6-43520)	63.8 (19.2)
L100PE13 F1 (4)	37.9 (9)	528 (34.4-42290)	63.3 (16.9)
C57BL6Sal F2 (5)	37.6 (6)	524 (37.3-40258)	69.2 (22.5)
L100PRoli F1 (5)	38.1 (6)	427 (28.8-39550)	67.3 (16.7)
C57BL6E13 F2 (5)	37.7 (5)	433 (37.0-39070)	66.4 (19.7)
Q31LBupro F2 (5)	37.8 (4)	484 (34.0-38169)	67.6 (21.2)
C57BL6E13 F1 (5)	37.5 (8)	412 (32.4-39894)	63.5 (21.2)
L100PE13 F2 (5)	37.1 (8)	477 (38.0-40084)	66.4 (21.7)
C67BL6Sal F2 (6)	37.5 (10)	692 (10.1-44156)	77.2 (20.1)
L100PSal F2 (6)	38.2 (12)	627 (12.3-47937)	76.2 (18.3)
C57BL6E13 M2 (6)	38.2 (7)	536 (14.6-47484)	69.1 (15.6)
L100PSal M1 (6)	38.2 (8)	589 (26.3-46734)	71.1 (17.5)
L100PRoli F2 (6)	38.2 (9)	589 (25.7-43120)	71.8 (18.2)
Q31LE13 F1 (6)	37.9 (9)	607 (27.9-46465)	72.4 (16.8)
C57BL6Sal F1(7)	37.2 (9)	514 (37.4-40969)	73.5 (15.6)
C57BL6Sal F1(7)	38.1 (8)	513 (35.8-43550)	67.9 (15.5)
C57BL6Sal F1(7)	37.9 (8)	551 (34.9-41587)	68.2 (17.1)
L100PE13 M1 (7)	38.2 (7)	505 (33.7-40032)	67.0 (14.7)
C57BL6E13 M1 (7)	38.4 (9)	477 (33.7-44449)	64.7 (16.9)
Q31LSal M1 (7)	36.9 (8)	509 (33.5-41563)	64.8 (16.1)

Table 3.7: Summary of raw expression data. Table includes pool ID and beadchip number (as in figure 3.4); average number of beads on each probe plus the minimum number of beads used to calculate probe intensity; the average probe intensity for each array plus the range of intensities recorded; and the background and noise recorded for each array. Replicate samples are shown in italics.

3.6 Quality Control

Multiple methods were employed to determine the quality of the microarray experiment, including both platform specific and sample specific controls as outlined below.

3.6.1 Illumina BeadChip Built in Controls

The Illumina BeadChip system has six in built control categories designed to monitor multiple aspects of the array experiment including sample labeling, hybridization to the array chip, quality of biological specimen and generation of probe signals. Quality control results from the Illumina built in controls in this experiment are shown in figure 3.7.

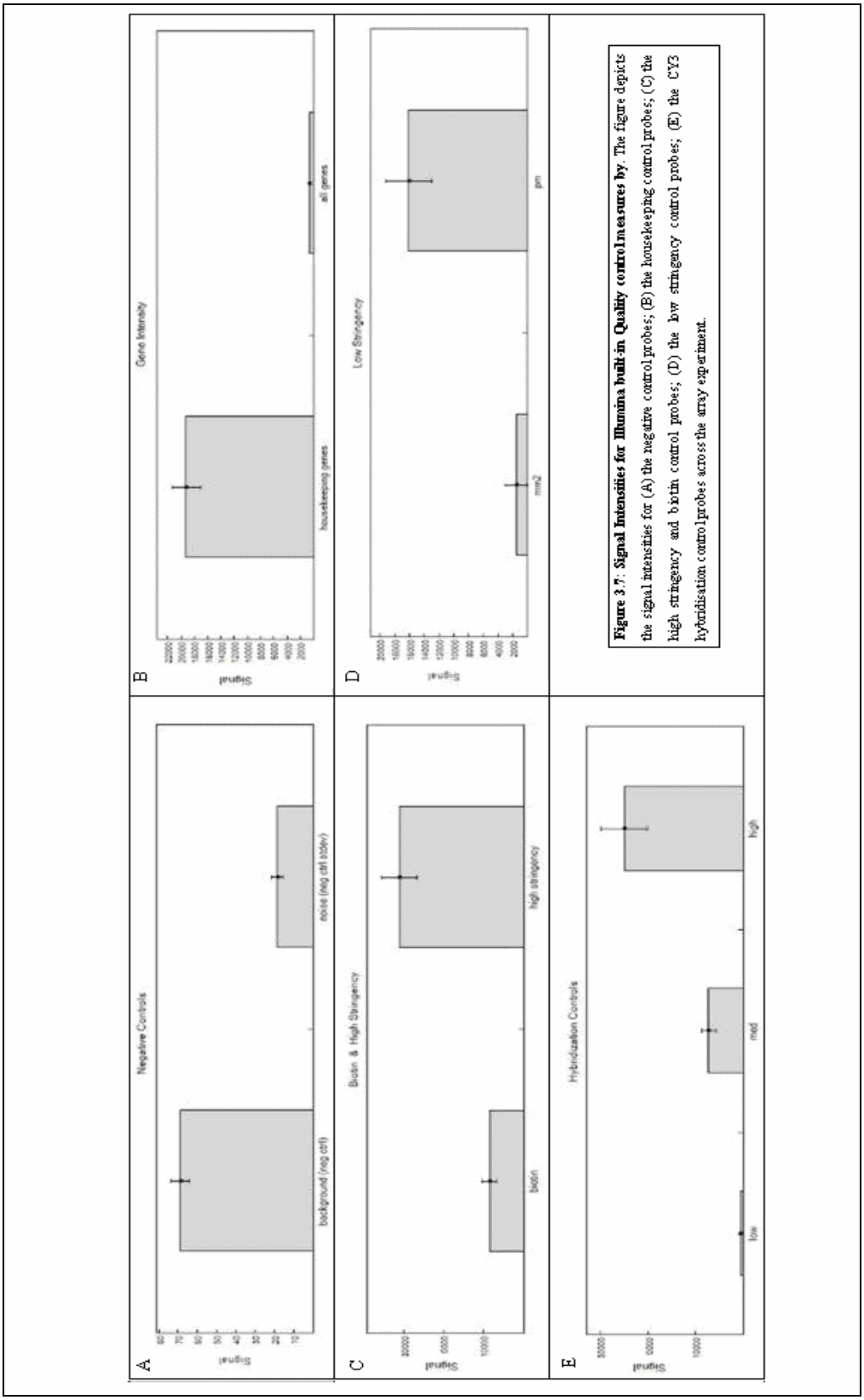

Figure 3.7: Signal Intensities for Illumina built-in Quality control measures by. The figure depicts the signal intensities for (A) the negative control probes; (B) the housekeeping control probes; (C) the high stringency and biotin control probes; (D) the low stringency control probes; (E) the CY3 hybridisation control probes across the array experiment.

The negative controls (figure 3.7:A) consisted of 1,602 probe sequences which were thermodynamically comparable to the gene specific probe sequences but were made up of random sequence without specificity in the genome. Signal obtained from these probes was therefore an artifact of the imaging system background, non-specific binding or cross-hybridisation. The mean signal produced from these negative probes defined the system background (~ 69) while the standard deviation of the probes determines the noise in the array (~19) as outlined previously in table 3.7. This is significant to $p<0.01$. This value is then used in further analysis to establish the limits of detection within the array to allow genes with expression values lower than the negative to be discarded as they are not expressed above background.

The housekeeping gene controls (gene intensity), as shown in figure 3.7 (B), consist of 24 probes targeting seven housekeeping genes designed to monitor the integrity of the samples on the array. As these genes are expressed well in all intact samples it is expected that in a high quality sample the bead intensity of the housekeeping genes will be considerably higher than the average of all genes. The bead intensity of the housekeeping genes in this experiment was ~18,900 compared to an average targeted gene intensity of ~555, indicating that the samples used were of high quality ($p<0.001$).

The high stringency control and biotin-labelled control are both shown in figure 3.7:C. The high stringency control consists of two probes with high G+C content which define the upper bounds of stringency and are expected to hybridise regardless of the level of stringency observed. Signal from these probes in the absence of signal from the other hybridization controls would indicate that the stringency was too high. This was not the case in this experiment and it was determined that overall stringency was acceptable.

The low stringency control consists of sixteen probes which determined the lower bounds of stringency (figure 3.7: D). The probes each contain two mismatch bases (mm2) which is compared to their perfect-match counterparts (pm2). If the mm2 approaches the pm2 the hybridization stringency is considered to be too low. In this case there is a significant difference between the mm2 and the pm2 values ($p<0.001$) indicating sufficient stringency of hybridization.

The hybridization controls (figure 3.7: E) consist of twelve Cy3-labelled oligonucleotide probes complimentary to those present in the hybridization buffer. This control is independent of sample quality and determines whether the hybridization step was successful during array preparation. The signal produced should be proportional to the concentration of the probe to indicate successful hybridization. As this was the case with the samples tested it was concluded that hybridization was successful.

The biotin-labelled control determines the success of secondary staining and signal generation due to the presence of biotin labeled oligonucleotides in the hybridization buffer. The three probes gave a positive hybridisation signal indicating successful secondary staining.

These primary quality control measures indicated that the RNA samples were of sufficient quality and hybridization to the BeadChips was successful in this experiment. However, I also analysed the control data from each array separately to compare quality between chips (figure 3.9), and found one chip appeared to be significantly different to the others tested.

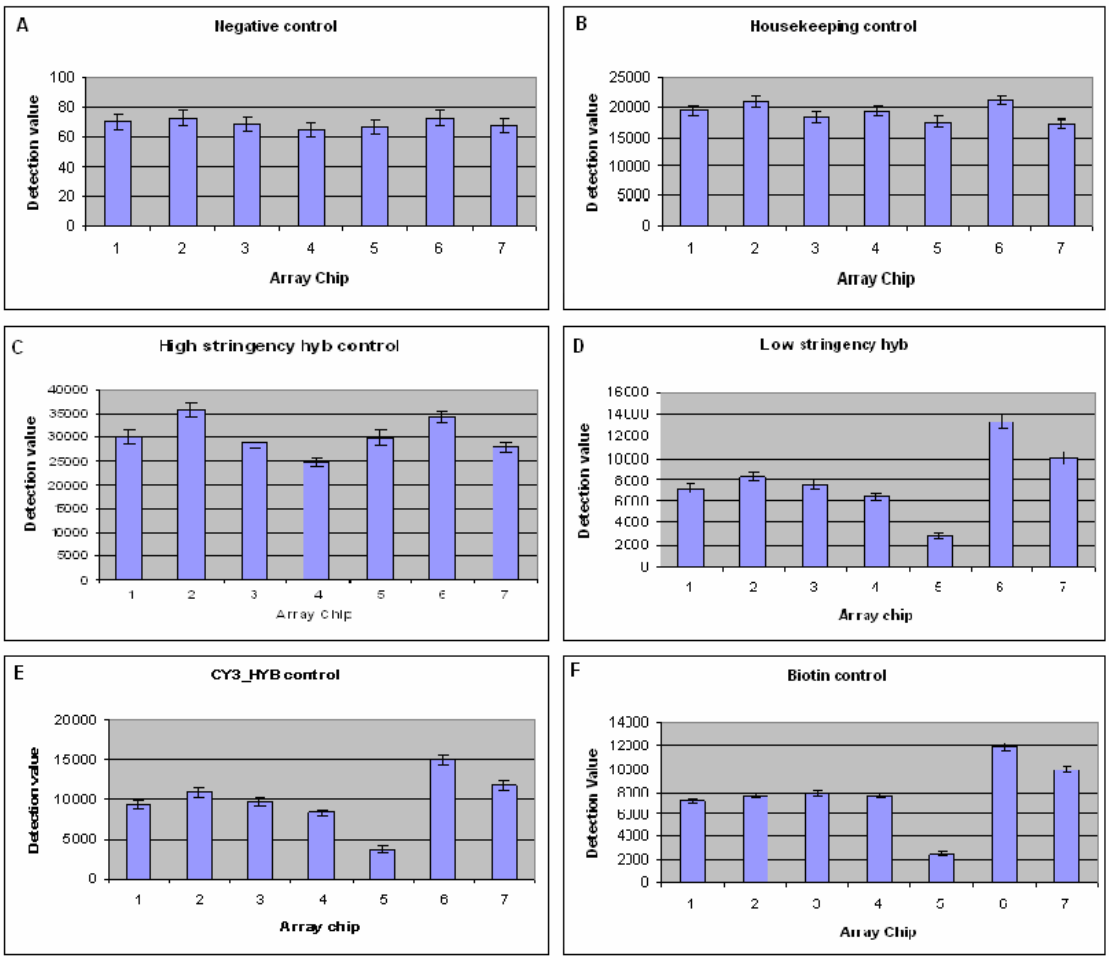

Figure 3.8: Signal Intensities for Illumina built-in Quality control measures by BeadChip. The figure depicts the signal intensities for (A) the negative control probes; (B) the housekeeping control probes; (C) the high stringency control probes; (D) the low stringency control probes; (E) the CY3 hybridisation control probes and (F) the biotin-labelled control probes for each individual BeadChip. BeadChip 5 has much lower probe signal intensities for low stringency hybridization, CY3 hybridisation and Biotin-labelled probes.

While BeadChip 5 passed the previous quality control measures when all chips were analysed together, further analysis revealed the CY3, low-stringency hybridization and biotin labeled control measures were considerably lower than those of the other BeadChips. After discussions with the Illumina support staff it was determined that part of the hybridization process for chip 5 was unsuccessful, probably due to the chip drying

out slightly during this step. As a result, samples that had been run on chip 5 were re-submitted to the WTCRF and hybridization of chip5 was repeated. The chip-by-chip control profiles with the new chip 5 data are shown in figure 3.9.

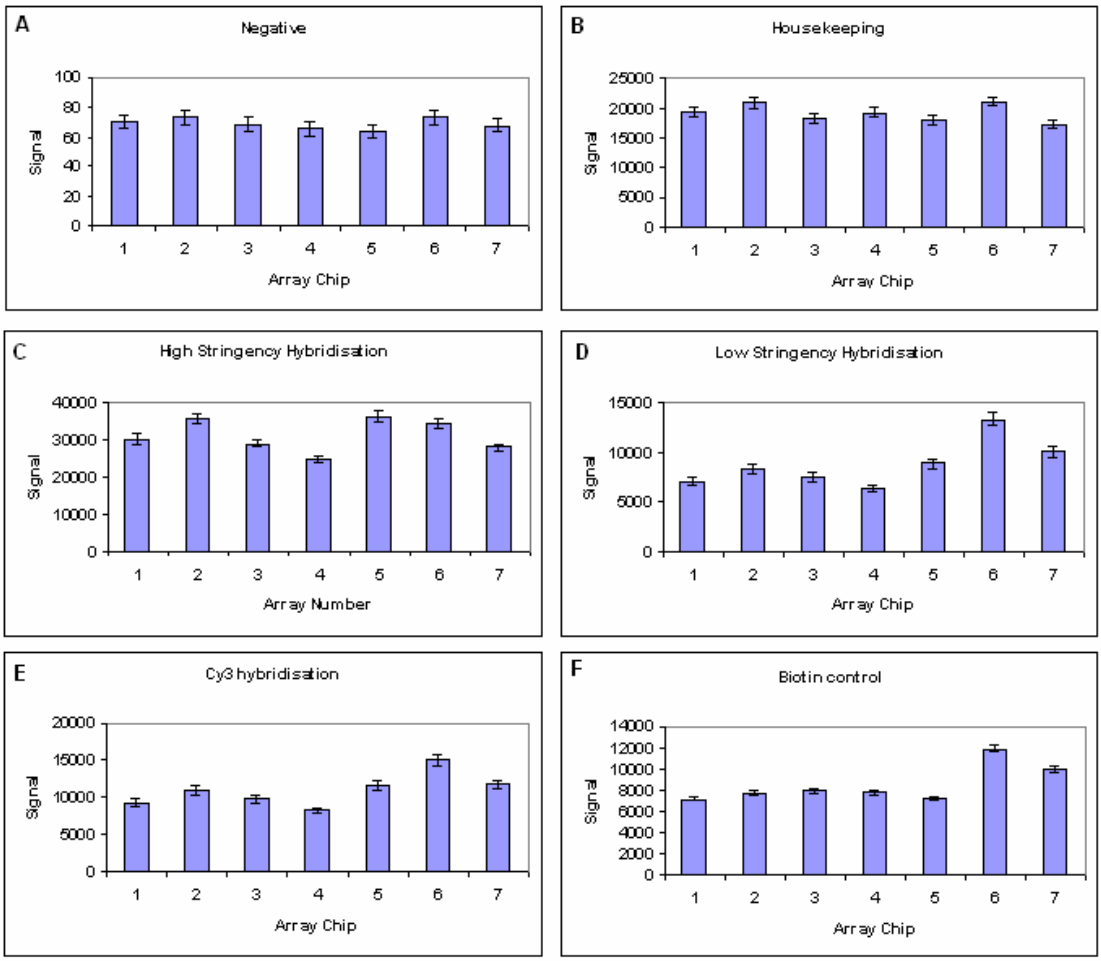

Figure 3.9: Signal Intensities for Illumina built-in Quality control measures by BeadChip post chip5 repeat. After the repeat run of BeadChip 5 the low stringency hybridization, CY3 hybridisation and biotin-labelled control probes produce comparable signal intensities to the other BeadChips, allowing more confident comparisons between chips to be made.

The repeat run of BeadChip5 gave results comparable to the control values previously observed for the other six BeadChips. From this point forward all analysis was carried out on the data from the repeated BeadChip5 run.

In summary, the initial analysis of the system controls did not indicate a problem with any BeadChip. Interquartile range analysis had not highlighted BeadChip5 as an outlier and primary quality control analysis was as expected in a successful array. Only the secondary analysis looking at each chip's controls individually showed a potential problem with one chip. I would therefore conclude that this step is vital in ensuring all system quality control measures are accurate.

3.6.2 Technical Replicates

To assess the reproducibility of the BeadChip platform and RNA preparation in this experiment technical and experimental replicates were included in this study. The reproducibility of the BeadArray platform was determined by the inclusion of the C57BL6 F1 pool on all BeadChips while the joint within-BeadChip and RNA preparation reproducibility was determined by the inclusion of three RNA preparations of the same sample (C57BL6 F2) on BeadChip 7.

Pairwise comparisons of the raw, untransformed C57BL6 F1 pool data (figure 3.10) revealed high correlation between the replicate samples across chips (Pearson's r^2 $\geq$0.972). Similarly high pairwise comparisons were observed between the RNA preparations of C57BL6 F2 (figure 3.11) on BeadChip7 (Pearson's $r^2 \geq 0.995$) indicating that variability due to batch and experimental effect were within acceptable ranges.

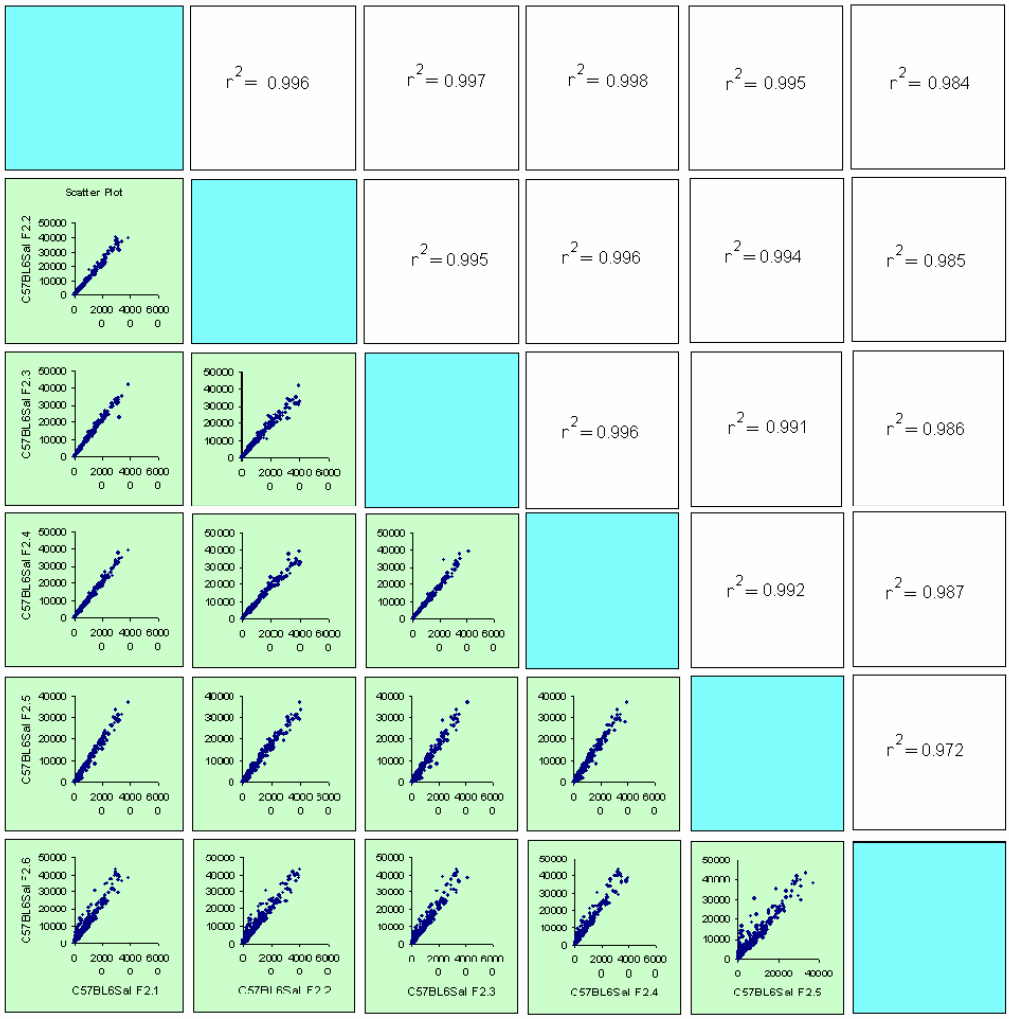

Figure 3.10: Pairwise Comparisons of raw data for C57BL6 F1 pool samples. Figure shows pairwise scatter plots of the log2-transformed intensities of the C57BL/6J chip replicates produced using the Bioconductor *lumi* package. Correlation estimates (Pearson's correlation coefficient r2) were calculated using the *cor* function in R.

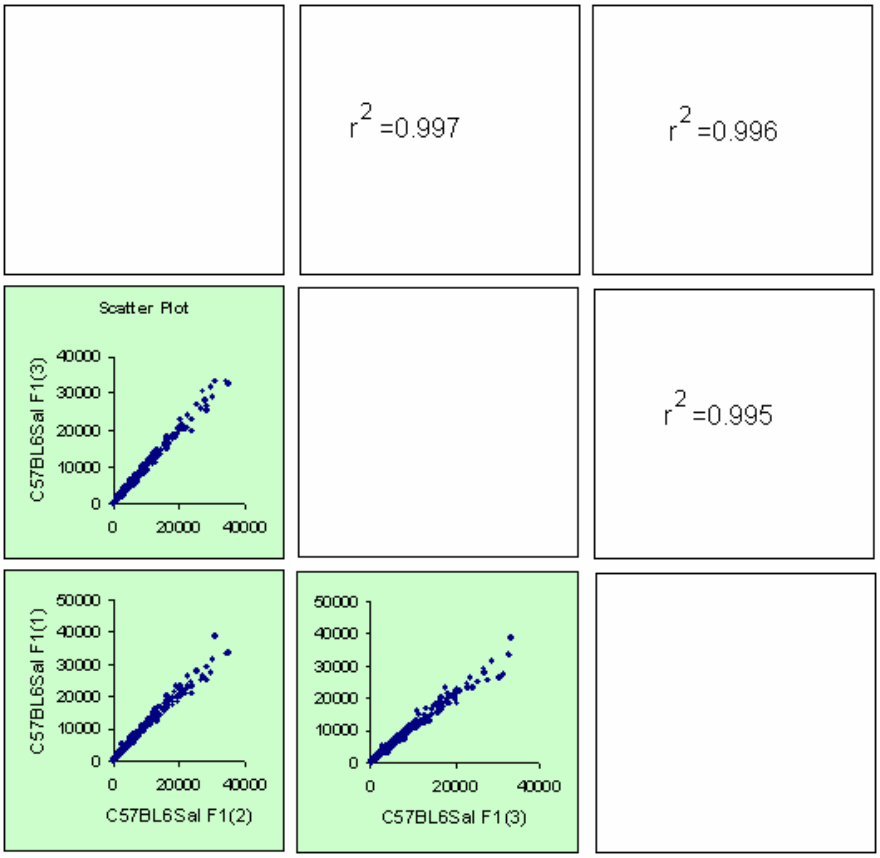

Figure 3.11: Pairwise comparisons of raw data for C57BL6 F2 pool samples. Figure shows pairwise scatter plots of the log2-transformed intensities of the C57BL/6J technical replicates produced using the Bioconductor *lumi* package. Correlation estimates (Pearson's correlation coefficient r2) were calculated using the *cor* function in R.

3.6.3 Quality Control using Cluster Analysis

Cluster analysis is a basic statistical approach which assigns samples into groups based on similarity. Hierarchical cluster analysis starts with single-member groups which are merged based on how similar their expression profiles are resulting in a hierarchical tree known as a dendrogram. Multiple methods of calculating hierarchical clusters exist which differ in the way they calculate similarity, and give different clusters [230] highlighting

different aspects of the data. The data was analysed by 'average correlation' using the freeware R with the **pvclust** application, that assesses robustness of the clusters using multiscale bootstrapping (1000 bootstraps). The cluster analysis was carried out on raw expression data to determine if technical or experimental artifacts affected clustering. It was expected that samples would not cluster into their respective biological group prior to data pre-processing and would show random spread across the dendrogram.

Figure 3.12 shows the average correlation dendrogram for the raw expression data. The pairwise distance (or similarity) was measured using the correlation method (1 pearsons correlation) and clusters joined using 'average linkage' which measures correlation distances across clusters. The bootstrapping results are represented by the approximate unbiased (AU) probability values, which indicate the frequency at which that particular cluster was observed over the 1000 bootstraps. An AU value of 100 indicates the cluster was produced on every occasion during bootstrapping.

At this stage the embryonic and adult samples are clearly separated into two distinct branches with a Pearsons correlation of around 0.75. This would be expected as the expression profiles of the embryonic samples would be markedly different to the adult samples. Within the two main branches there was a maximum distance between two samples of ~0.025, indicating a Pearsons correlation of $\geq$0.975 which suggests low overall variability even before pre-processing and data stabilization. The technical replicates (C57BL6 F1) cluster together within the adult branch with an AU value of 100, and the chip replicates (C57BL6 F2 hybridised to each chip) also cluster together with the exception of the sample on chip 6 which falls closer to the technical replicate clusters. All other groups appear evenly spaced throughout the dendrograph with no observable clustering due to chip number or gender (figure 3.12).

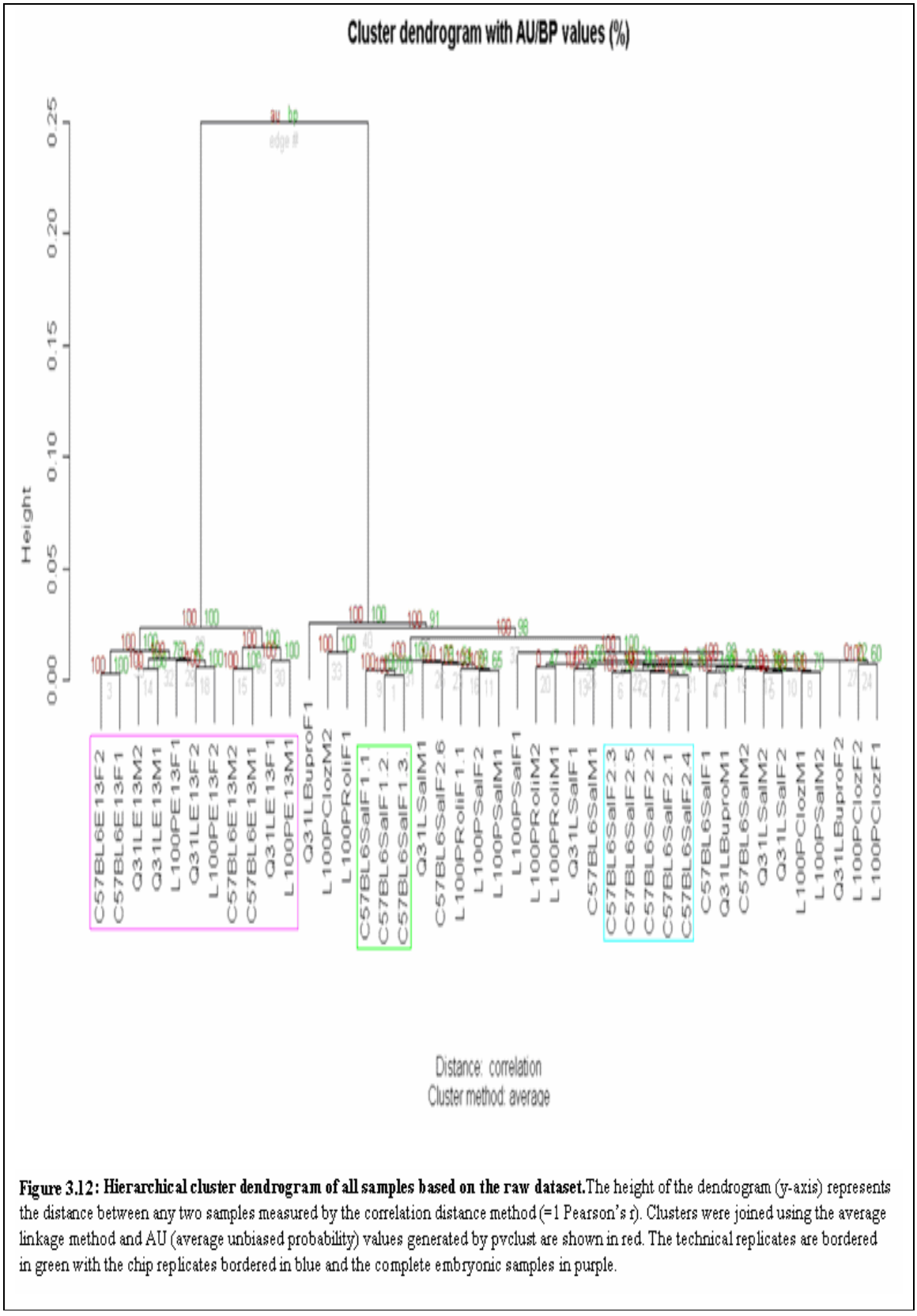

Figure 3.12: Hierarchical cluster dendrogram of all samples based on the raw dataset. The height of the dendrogram (y-axis) represents the distance between any two samples measured by the correlation distance method (=1 Pearson's r). Clusters were joined using the average linkage method and AU (average unbiased probability) values generated by pvclust are shown in red. The technical replicates are bordered in green with the chip replicates bordered in blue and the complete embryonic samples in purple.

3.7 Data Pre-processing

The Illumina microarray platform is unique in having a high redundancy of randomly positioned beads, which reduces positional bias, increases robustness of hybridization intensity and measurement error estimates across each probe. As multiple samples are hybridized to each chip, hybridization batch effects are also reduced. These features mean pre-processing of the data prior to differential expression analysis should be carefully considered to take advantage of this redundancy.

Three levels of pre-processing were carried out on the raw data in this study. Firstly the variance within the data was stabilized, secondly the stabilized data was normalized, and finally the stabilized and normalized data was filtered to first remove any probes not expressed above background and then to remove any probes that were not expressed at a level considered biologically significant.

3.7.1 Variance Stabilisation

Microarray data has a tendency to be highly variable, with high intensity data displaying larger variance. Inconsistency in variance (heteroskadacity) is a violation of many parametric statistic tests, such as analysis of variance (ANOVA), which assumes constant equal variance. To stabilize the variance in this dataset the variance stabilizing function (lumi-vst) within the Illumina specific Bioconductor **lumi** package (www.basic.northwestern.edu/projects/lumi) was used. Again, previous studies within the laboratory had shown this to be more reliable than standard $\log_2$-transformation, as the lumi-vst function directly models the mean-variance relationship in the array by taking account of the within-array bead redundancy to generate an optimal transformation parameter. After stabilization of the data using the lumi-vst function, a generalized log transformation was then performed to give appropriate complete stabilization of the data and more manageable data values for further analysis.

3.7.2 Normalisation

The second step in pre-processing is to normalize the data to remove the systematic variation caused by non-biological sources such as RNA quality, hybridization and scanning. Again the Bioconductor package lumi was utilized in this step using the lumi-rsn function which was originally created to be used in conjunction with the lumi-vst function from the previous step (www.bioconductor.org/packages/2.1/bioc/vignettes/lumi/inst/doc/Bioc2007_lumi_presentation.pdf). Due to the size of the dataset it was necessary to process each chip individually. This was possible as each array was normalized against a target array which was most similar to the mean of all samples. The mean reported signal intensity across the whole array was ~555. With a mean signal intensity of ~551, the technical replicate C57BL6 F1 (1) was the closest to the overall mean and was used as the target array for normalization. All samples on the array were normalized to this array using the entire probe-set.

Heirarchical cluster analysis was carried out on the stabilized and normalized data (figure 3.13) showing a reduction in overall variability across the array. At this stage the embryonic and adult samples are still clearly separated into two distinct branches with a Pearsons correlation of around 0.85. This would be expected as the expression profiles of the embryonic samples would be markedly different to the adult samples. Within the two main branches there was a maximum distance between two samples of ~0.015, indicating a Pearsons correlation of ≥0.985. The technical replicates (C57BL6 F1) cluster together within the adult branch with an AU value of 100, and the chip replicates (C57BL6 F2 hybridised to each chip) also now all cluster together. All other groups still appear evenly spaced throughout the dendrograph although with greater correlation than seen previously, with no observed clustering due to chip number or gender. Pairwise correlation of the technical replicates increased to r^2 ~0.996 and chip replicates to ~0.991. Formal testing of the data using analysis of variance (ANOVA) with 41 degrees-of-freedom showed no overall significant variability of the data indicating normalization was successful ($F=5.23$, $p=0.0006$).

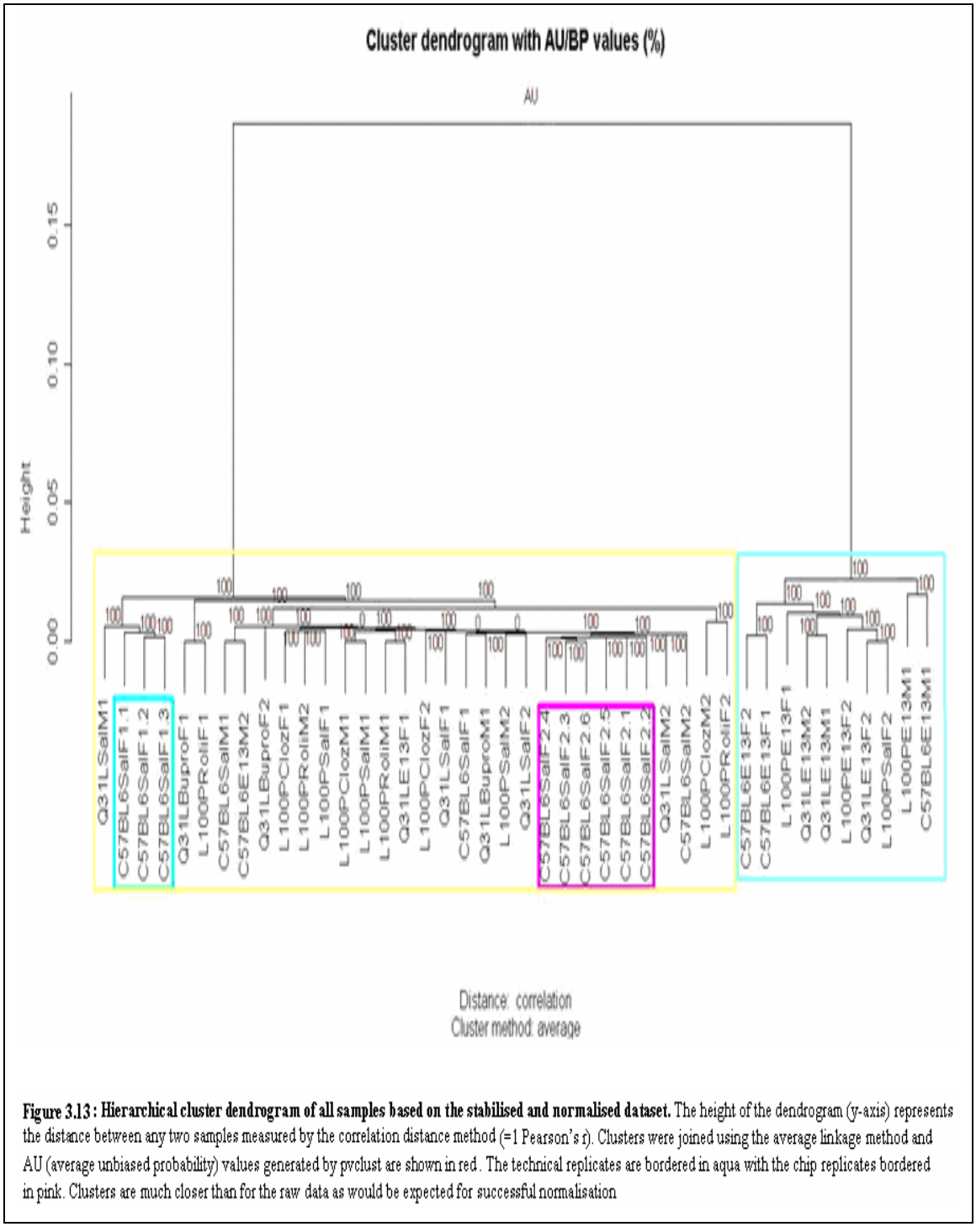

Figure 3.13 : Hierarchical cluster dendrogram of all samples based on the stabilised and normalised dataset. The height of the dendrogram (y-axis) represents the distance between any two samples measured by the correlation distance method (=1 Pearson's r). Clusters were joined using the average linkage method and AU (average unbiased probability) values generated by pvclust are shown in red. The technical replicates are bordered in aqua with the chip replicates bordered in pink. Clusters are much closer than for the raw data as would be expected for successful normalisation

As global background correction would have resulted in negative signals which would have affected the pre-processing steps this was not performed on the datasets. Instead the data was filtered to include only the probes that were expressed above background in each sample as described in section 3.7.3 below.

3.7.3 Data Filtering

Data filtering was used to reduce the multiple testing burden and increase sensitivity of the microarray analysis. The data was filtered in two steps; the first removed probes that were not expressed above background, and the second selected only the probes with biologically relevant differential expression levels.

The detection limits of the array experiment are automatically generated from the background and noise measurements which are estimated from the signal intensities produced by the built-in negative controls. The detection P-value for each probe is calculated for individual and multiple samples using non-parametric methods that rank the probe signal relative to the negative control. This can be interpreted as the probability that the probe signal is no greater than background, relative to the noise in the data. The higher the intensity signal, the lower the probability that it is generated by non-specific sources. The detection P-value was calculated for each probe in each of the nine biological groups (C57BL6Sal, L100PSal, Q31LSal, L100PCloz, L100PRoli, Q31LBupro, C57BL6E13, L100PE13 and Q31LE13) and only those probes with a detection P-value of less than 0.05 in at least one of the nine groups were retained for further analysis. This allowed probes to be expressed in one group but not in another. Of the 46,644 successful probes (corresponding to 34,492 genes) on the Mouse Ref-6 BeadChip, 21,869 (~47%) probes met this threshold. Estimates by others state that most tissues only express 30-40% of genes [231] so this is slightly higher than would be expected. Figure 3.14 illustrates the final cluster dendrogram after the data had been filtered. There is very little difference between this and the previous dendrogram for the stabilized and normalized data. This was unexpected as filtering of probes not significantly expressed above background should increase the bootstrapping AU values

indicating stronger support for the observed clusters. However, the previous cluster dendrogram gave AU values all close to or equal to the maximum of 100 which is likely to be the reason that no difference is seen in this new filtered dataset.

In the second filtering step only probes which showed potentially significant differential expression when compared to the relevant control group were retained. This has previously been shown that the most reproducible and reliable method of determining differential expression in microarray analysis is based on a rank of fold-change and non-stringent statistical significance (P-value) cut-offs [225]. As the Illumina platform offers the sensitivity to detect down to a 1.3 fold change (www.illumina.com/downloads/Mouse6-8v2Datasheet.pdf) this was used as the cut-off for this second filtering step.

Fold change was calculated by taking the anti-$\log_2$ of the difference between the average normalized values of the two groups being compared at each probe. This fold-change filter reduced the number of probes to be tested from 21,869 to 1266 in the L100PSal vs C57BL6Sal comparison, 26 in the Q31LSal vs C57BL6Sal comparison, 996 in the L100PCloz vs L100P Sal comparison, 1313 in the L100PRoli vs L100PSal comparison, 251 in the Q31LBupro vs Q31LSal comparison, 2005 in the L100PE13 vs C57BL6E13 comparison and 622 in the Q31LE13 vs C57BL6E13 comparison (Figure 3.15).

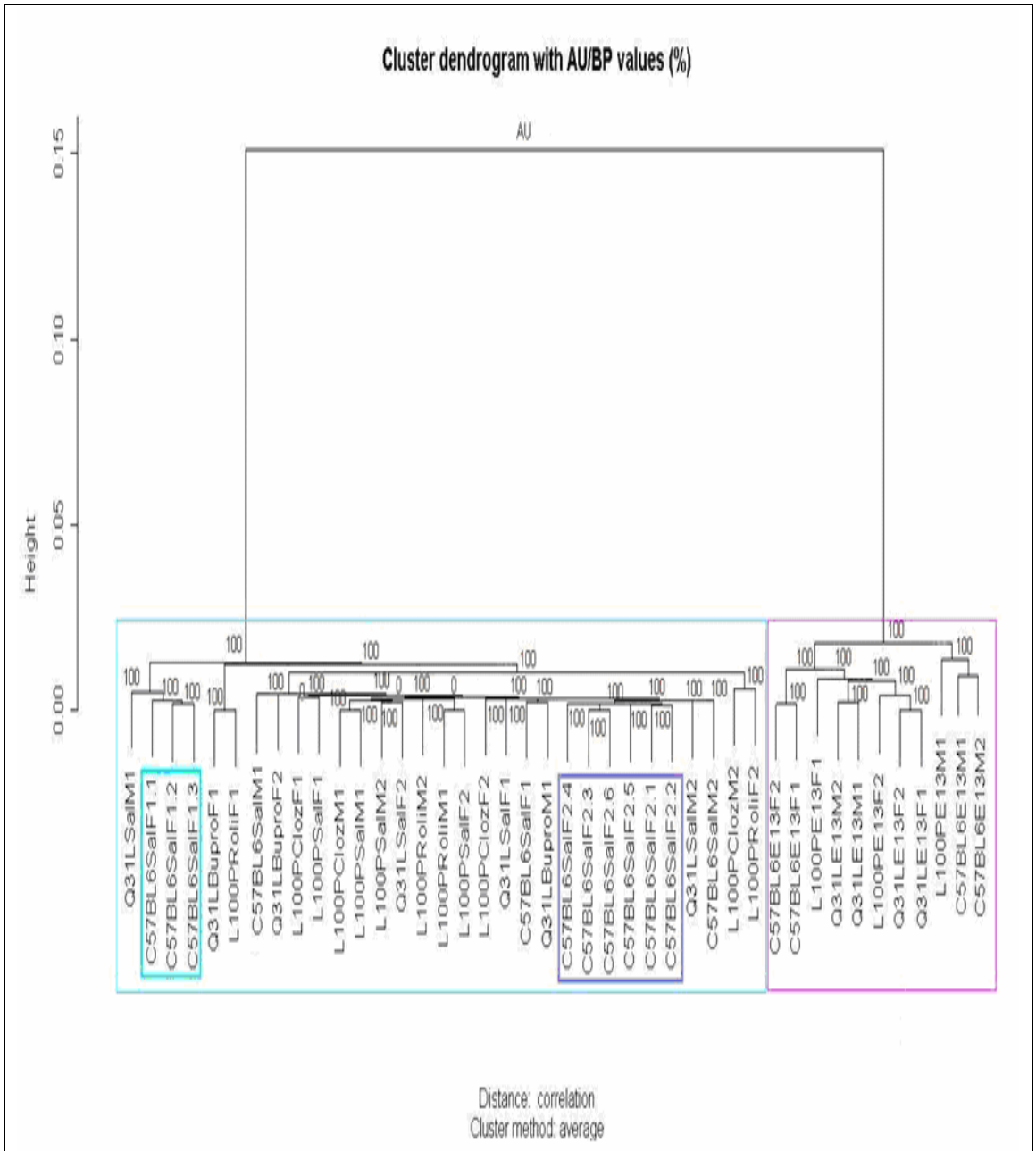

Figure 3.14: Hierarchical cluster dendrogram of all samples based on the filtered, transformed and stabilised dataset. The height of the dendrogram (y-axis) represents the distance between any two samples measured by the correlation distance method (=1 Pearson's r). Clusters were joined using the average linkage method and AU (average unbiased probability) values generated by pvclust are shown in red. The technical replicates are bordered in aqua with the chip replicates bordered in purple and the complete embryonic samples in pink.

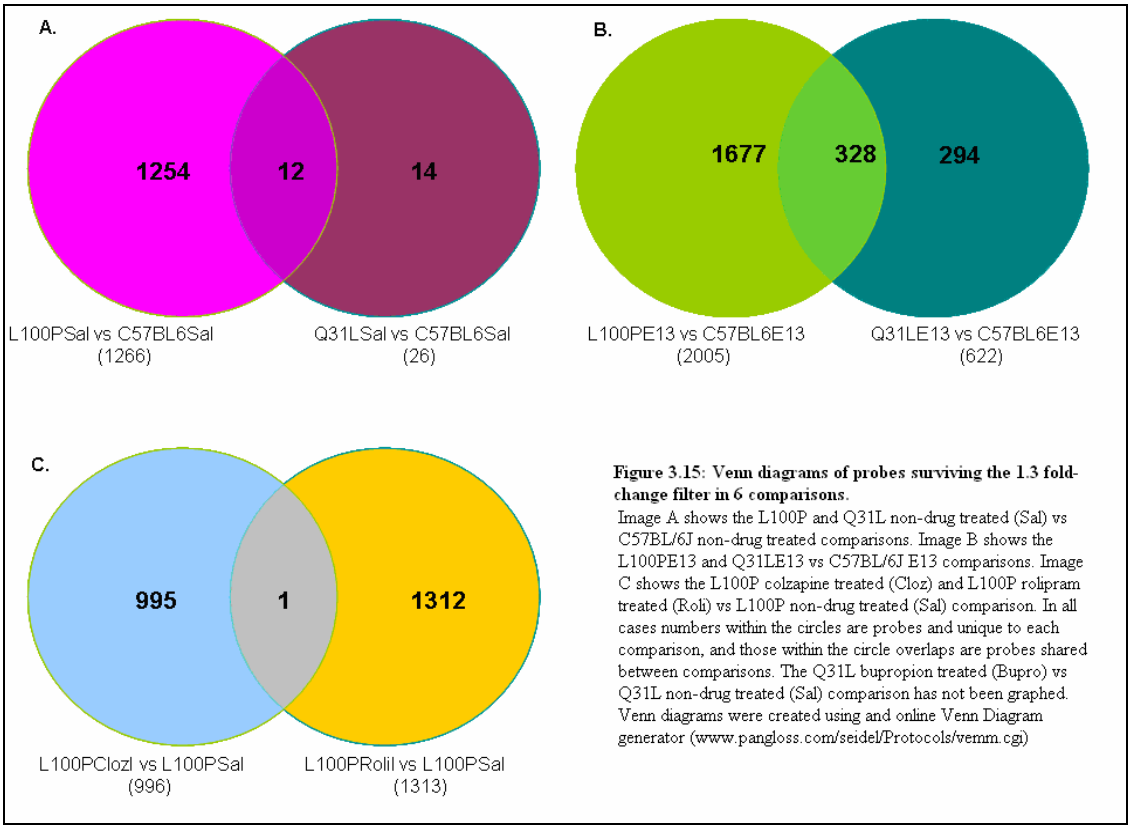

Figure 3.15: Venn diagrams of probes surviving the 1.3 fold-change filter in 6 comparisons.
Image A shows the L100P and Q31L non-drug treated (Sal) vs C57BL/6J non-drug treated comparisons. Image B shows the L100PE13 and Q31LE13 vs C57BL/6J E13 comparisons. Image C shows the L100P colzapine treated (Cloz) and L100P rolipram treated (Roli) vs L100P non-drug treated (Sal) comparison. In all cases numbers within the circles are probes and unique to each comparison, and those within the circle overlaps are probes shared between comparisons. The Q31L bupropion treated (Bupro) vs Q31L non-drug treated (Sal) comparison has not been graphed. Venn diagrams were created using and online Venn Diagram generator (www.pangloss.com/seidel/Protocols/venm.cgi)

3.8 Conclusions

This chapter has described the experimental preparation and pre-processing analysis of a genome wide expression study carried out on two ENU mutant mouse lines which display distinct endophenotypes categorised as 'schizophrenic-like' and 'depressive-like' using the Illumina BeadArray platform.

The behavioural study carried out prior to the collection of samples for the microarray was designed to replicate the results obtained by Clapcote *et al* (2007) [197] to ensure the mutation continued to carry the behavioural phenotype previously described. As described, the animals used for this experiment were backcrossed for a further 3 generations than those used in the previously published study and the pool of mice available was substantially affected by a spontaneous mutation within the breeding

programme which required a mass cull of animals prior to the breeding of those to be used in this experiment. This resulted in a much reduced gene pool available, and the requirement to cross homozygous animals to generate the number of animals required, which further compounded the bottleneck effect. This also made it impossible to use wild-type littermates as a control as no heterozygous animals were used in the crosses. As the ENU mutant lines were established on a C57BL/6J genetic background it was determined that these animals would be the closest genetically to wild-type littermates.

The PrePulse inhibition analysis carried out by myself did not replicate the study previously published by Clapcote *et al* using the ENU mutant mouse lines. There are a number of reasons which may explain this. It is possible that the further backcrossing reduced the effect of the mutation on the behaviour phenotype previously observed. This is unlikely however, as additional, more recent, studies (Tatiana Lipina, personal communication) have replicated the results previously published, and personal observations of the animals brain physiology and general behaviour concur with those reported previously. Subsequent literature searches combined with results obtained in this experiment show the C57BL/6J mice may not have been the ideal control for this experiment. Studies of PPI in rodents have been going on for many years as researchers attempt to find the genetic basis for such behaviours. Paylor and Crawley (1997) [232] studied the differences in PPI in commercially available inbred mouse strains. One strain used in their study was the C57BL/6J strain from Jackson Laboratories as used in this study. They looked at mice between the ages and 9 and 14 weeks over 5 auditory prepulse stimuli. Interestingly, the C57BL/6J mice gave the lowest PPI of all the mice they tested[232]. Two of the decibel levels used by Paylor and Crawley [232] can be compared to those used in the current study and give an average of 20% PPI compared to the 23% PPI observed in the current study. It is also important to note that the original Clapcote [197] experiment grouped the wild-type littermates from both mutants in one ‘control’ group so any differences which may appear between wild-types littermates will be hidden within the one group. While the C57BL6 mice were specifically sourced from the Jax substrain for the behavioural study, to be as close as possible to the original backcross strain used when establishing the ENU mutant colonies, the results obtained would suggest that they were not an adequate control, and there may be non-target

mutations in the wild-type littermates which may be responsible for the differences observed. These could either be secondary ENU mutations that co-segregated with the *Disc1* mutation, or sections of DBA genome that remained, despite backcrossing, around the *Disc1* gene. This is discussed further in Chapter 8 and any future work should review the validity of non-littermate controls. It should also be noted that while comparison between PPI deficits in rodents and humans are, for the most part consistent, some drug effects have been shown to give differing results across species[220]. Prepulse inhibition is regarded as a reliable endophenotype of major mental illness (and schizophrenia in particular) but care must be taken over the conclusions drawn from our results if we wish to relate our mouse model to human disease.

Despite the limitations described above, it was determined that the full microarray experiment should continue as any genes which were 'flagged' as being differentially expressed in the comparisons with the C57BL/6J mice would be subsequently tested against wild-type littermates after the array as it would be possible to re-establish a complete ENU mutant mouse colony within this time.

The aim of the microarray experiment was to determine which genes were differentially expressed on the basis of genotype, drug treatment and developmental stage (embryonic and adult). A total of seven comparisons were made across nine biological groups. The tissues used for the array analysis were taken from the animals used in the behavioural study immediately after completion of the PrePulse inhibition cycle. All animals that underwent behavioural testing were sacrificed and RNA extracted from hippocampal tissues, however only the groups outlined previously were included in the array. Abnormalities in hippocampal structure, activation, organization of neurons and synapse function have been well documented in human schizophrenia patients and other mouse models [207, 214, 233, 234]. As there are working memory deficits often associated with schizophrenia [208], it is clear there is substantial hippocampal involvement in the disorder. The use of a single brain region also removed any confounding effects of compensation in other brain regions, which may have diluted the result. Financial constraints meant that only the drug treated groups which had shown a marked behavioural response in the previously published study were included, as these were

more likely to give a distinct genetic profile for comparison. The RNA from the other groups was retained to be used for future experiments and to determine the importance of some of the results obtained (as described in Chapter 5). The sample preparation, quality control and pre-processing methods used had been previously tested by other researchers within the laboratory and were considered of good standard with reliable output for differential expression analysis. Biotin labeling and hybridization controls indicated the RNA used was of good quality, increasing the likelihood of a reliable output from the array.

The main limitation of this study was the requirement to use the C57BL/6J mouse as a wild-type control, particularly with the resulting PrePulse inhibition result, and if the experiment were to be repeated it would be advisable to use wild-type littermates as controls. The use of these animals was however justified by the fact the ENU mutants had been backcrossed for 10 generations and would be considered as close to the C57BL/6J genotype as possible, and further analysis would include wild-type littermates to validate any results. As RNA was available for other biological groups not included on the array it would also have been interesting to compare the drug treated C57BL/6J animals with the drug treated ENU mutant samples to potentially determine whether the drug treatments affected the ENU mutants in a different manner to their 'wild-type' controls.

The next stage of analysis will describe differential expression profiles and potential biological significance of the results obtained from this primary analysis.

Chapter 4
Differential Expression Analysis of the Disc1 ENU Mutant Mouse Microarray

4. Differential Expression Analysis, Primary Annotations and Biological Relevance

4.1 Introduction

Analysis of differential expression between mutant lines provides insight into potential genes and pathways involved in the phenotypes expressed in these animals. By comparing these genes to those previously published in human studies we can identify genes and pathways involved in major mental illness and potentially determine targets for further study or new drug treatments. This chapter describes the process of differential expression analysis in the ENU mouse mutants including pathway analysis and overlaps with previously published human studies.

4.2 Differentially Expressed Genes

The array probes that survived the fold-change filter described in chapter 3 were tested for differential expression using the Analysis of Variance model (ANOVA) written in the R package (MAANOVA) which is specifically designed to be used with microarray data [235]. The two main effects tested were genotype and drug treatment. Tests of differential expression were carried out using the MAANOVA *matest* function using the Fs test statistic, which makes no assumption about the distribution of variance but estimates it based on the information across all probes in the dataset. This has previously been shown to be more powerful than other F statistics available in the package for detecting differential expression [236] because of the employment of the James-Stein shrinkage concept. This concept uses the information across all probes, rather than probe specific estimators, to gain a better estimate of variance within the dataset. This increases the power of the test thus giving a higher reward in small sample sizes than standard F statistics. Table 4.1 shows the number of genes found to be differentially expressed in each group comparison (Fs p-value ≤ 0.05) based on genotype or drug treatment where appropriate.

For each probe tested I also calculated the false discovery rate (FDR), using the *adj* p-value function in MAANOVA, to determine the stringency of the 0.05 cut-off. The FDR value determines the expected proportion of false positives when determining significance of each probe. Table 4.1 shows the mean FDR for each comparison based on the *adj* p-value for each probe called as significant within that comparison. The maximum FDR for any comparison was 0.35 suggesting a moderate probability of any probe being a false positive. As genes were to be subjected to further rounds of validation this was deemed to be acceptable.

Comparison	Probes	Significant (%)	Mean FDR (maximun)
L100Padult vs C57BL6adult	1266	531 (41.9)	0.26 (0.3)
Q31Ladult vs C57BL6adult	26	9 (34.6)	0.21 (0.35)
L100PE13.5 vs C57BL6E13.5	2005	33 (1.6)	0.19 (0.22)
Q31LE13.5 vs C57BL6E13.5	622	15 (2.4)	0.22 (0.24)
L100Padult vs L100PRoli	1313	44 (3.4)	0.13 (0.18)
L100Padult vs L100PCloz	996	3 (3.0)	0.20 (0.23)
Q31Ladult vs Q31LBupro	251	103 (41.0)	0.18 (0.21)

Table 4.1 Summary of probes differentially expressed by genotype of drug treatment. Column 1 shows the pair-wise comparison made while column 2 gives the number of probes that passed the 1.3 fold-change filter. Column 3 shows the number and percentage of probes that were significantly differentially expressed (Fs p-value ≤ 0.05). The final column gives the mean and maximum false discovery rate (FDR) observed in the list of significantly differentially expressed probes.

Some genes showed differential expression in more than one pair-wise comparison (figure 4.1)

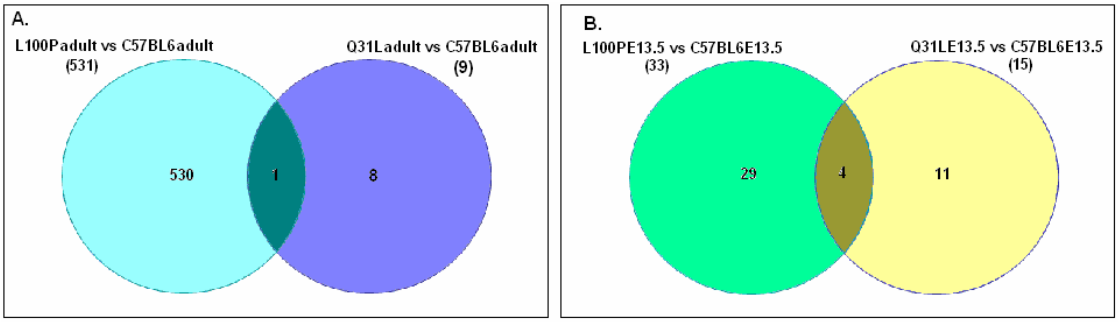

Figure 4.1 Venn diagram of significant probes in the genotype comparisons. Pane A shows significant probes in both adult comparisons while pane B shows significant probes in both embryonic comparisons. The total number of differentially expressed genes are shown in parenthesis. Within the circles of the Venn diagram are the number of genes unique to each comparison and, where the circles overlap, the number common to the comparisons.

In addition, eight genes that were found to be differentially expressed in the mutant vs wild-type comparisons were also differentially expressed when the mutant, non-drug treated (saline) animals were compared to their drug treated counterparts. When the drug treated animals were compared to the C57BL/6J non drug treated adults for expression at these genes no significant difference was observed, suggesting the drug treatment has corrected the expression level back to that of the control animal. Seven of these genes were in the L100Padult (saline) vs L100PRoli (rolipram) comparison while one gene was in the Q31Ladult (saline) vs Q31LBupro (bupropion) comparison. No correction of differential expression was observed in the L100PCloz (clozapine) group.

4.2.2 Differentially expressed genes: primary annotation and gene ontology

Genes that were shown to be differentially expressed were annotated using data provided by Illumina. Each probe sequence was also run through NGBI Gene to determine its location on the mouse genome and to check whether this corresponded to the gene name provided by Illumina. The official gene name, symbol, accession number, GI number and gene ontology were noted for each gene within this search.

Gene ontology (GO) analysis was performed to determine if the lists of significantly differentially expressed genes showed enrichment for any particular GO category relating to molecular function, biological processes or cell components. This was carried out using the web based GOTree machine (GOTM) (http://bioinfo.vanderbilt.edu/gotm) by comparing the list of differentially expressed genes to those expressed above background in the relevant groups. GO categories are arranged in a hierarchical structure of common, controlled vocabulary used to describe the roles of genes and gene products of different species. Only GO enrichments that are statistically significant ($P<0.01$, as determined by the hypergeometric test) are reported by GOTM.

The list of statistically significantly differentially expressed genes was then compared to gene lists from recently published studies. Three of the studies used for comparison [83, 84, 217] described copy number variants in cases and controls of schizophrenia and other major mental illness. A further two, unpublished, microarray studies identified genes dysregulated in two large families with a history of mental illness. The first used lymphoblastoid cell lines from a large Scottish family with a balanced translocation (t1:11) that formed the original sample set used to identify the DISC1 mutation and the motivation for characterising the two Disc1 mutants used in this study [8, 56]. This family has a history of schizophrenia and schizoaffective disorders spanning multiple generations. The second used lymphoblastoid cell lines from the F22 family which has a history of major depressive disorder and provides a dataset of both affected and unaffected individuals with married in controls[237]. Lastly, the DISC1 interactome [132] was used to identify dysregulated genes in this dataset that are known to interact with Disc1.

Finally, pathway analysis was carried out on the lists of differentially expressed genes using the web-based software Ingenuity Pathway Analysis (IPA). All probes expressed above background in each comparison were loaded into the system regardless of fold-change or p-value. These were then set as the main observations for determining cut-offs (ie, only include genes with $FC>1.3$ and $p\text{-val}<0.05$) so any pathways which were created by the system would highlight those genes with significant differential expression and

would also show (by shading) those genes expressed above background in the sample group. Any genes in the pathway which were not expressed above background in the sample group would be left unshaded. IPA then assigns a significance score to each network (as described in section 2.7.2) which indicates the likelihood that the assembly of focus genes within a network could be explained by random chance.

After these stages had been carried out all genes identified by the array analysis were scored on the basis of gene ontology, overlaps with previous studies and presence or absence in high scoring networks, to determine which genes would be carried forward for further analysis. The following sections describe the results of this process for each pair-wise comparison from the array analysis.

4.2.3 Differential Expression between the L100P adult ENU mutant and the C57BL/6J adult controls

Of the 1266 probes that passed the pre-processing filters in the L100Padult group, 531 showed differential expression compared to the C57BL/6J control group (table 4.3). 261 of these genes were over-expressed and 270 under-expressed in the L100P adult group. Four of these genes (*Akap9, Gnb1, Ppm1e and Smarce1*) are known interactors of Disc1 (Camargo et al 2006 [132]). With 127 genes in the Disc1 interactome, hypergeometric probability suggests this is a significant result (p=0.003). *Akap9, Gnb1* and *Smarce1* were over-expressed and *Ppm1e* under-expressed in the L100Padult sample group relative to the C57BL/6J controls.

TargetID	Full Gene Name	Accession Number	ProbeID	GI numbers	C57Sal MEAN expression	L100PSal MEAN expression	Difference	anti-log	fold change	Pval	GO Terms
WDFY1	WD repeat and FYVE domain containing 1	NM_001111279	1580463	141803291	8.167410917	8.753826628	-0.586415711	0.665995479	1.501511694	5.51E-11	metal ion binding
D930049F02RIK	RIKEN full-length enriched library, clone:D930049F02	AK086749	106400020		8.468013771	7.892311092	0.575702679	1.490403202	-1.490403202	5.26E-09	N/A
ENTPD4	ectonucleoside triphosphate diphosphohydrolase 4	NM_026174	7040288	142349997	10.62385611	9.958484121	0.665371992	1.585977157	-1.585977157	6.48E-09	integral to membrane, hydrolase activity, calcium ion binding
S100A8	S100 calcium binding protein A8 (calgranulin A)	NM_013650	70112	113930764	7.942089005	8.71456188	-0.772472875	0.585413177	1.708195235	2.92E-08	calcium ion binding, chemotaxis
MSR2	Fc receptor-like S, scavenger receptor	NM_030707.3	60204	148747157	9.409125746	9.024951496	0.38427425	1.305203038	-1.305203038	1.48E-07	scavenger receptor activity, extracellular
ABHD4	abhydrolase domain containing 4	NM_134076	6760450	19527301	10.56568473	9.940219202	0.625465532	1.54270855	-1.54270855	6.98E-07	aromatic compound metabolic process, lipid catabolic process, proteolysis
SCL0001284.1_18 (Mxra7)	matrix-remodelling associated 7	AK021409.1	106400100	12862292	8.170850589	8.651392671	-0.480542082	0.716708276	1.395267829	3.28E-06	N/A
SLC4A7	solute carrier family 4, sodium bicarbonate cotransporter, member 7	NM_001033270	2370056	117320528	9.32144757	9.808798836	-0.487351266	0.71333355	1.401868733	4.25E-06	integral to membrane, ion transport, sodium ion binding
PSMD1	proteasome (prosome, macropain) 26S subunit, non-ATPase, 1	NM_027357	102470156	142351621	12.6837714	13.11512968	-0.431358281	0.741563283	1.348502579	9.00E-06	cytosol, protein complex, binding
EXOC4	exocyst complex component 4	NM_009148	2470253	118130512	9.476114422	9.883417598	-0.407303176	0.754031565	1.326204427	1.48E-05	vesicle docking during exocytosis, protein transport
CHAC1	ChaC, cation transport regulator-like 1 (E. coli)	NM_026929	7100725	142385952	9.573771447	9.009318637	0.56445281	1.478826509	-1.478826509	2.99E-05	N/A
IMPAD1	inositol monophosphatase domain containing 1	NM_177730	2690594	118130493	9.573445011	10.03005872	-0.456613708	0.728694644	1.372316934	3.27E-05	integral to membrane, metal ion binding, inositol or phosphatidylinositol phosphatase activity
SENP1	SUMO1/sentrin specific peptidase 1	NM_144851	6100537	142353113	8.161569577	8.54307039	-0.381500813	0.767638613	1.302696325	5.96E-05	proteolysis, organismal development, hydrolase activity, peptidase activity
BEX2	brain expressed X-linked 2	NM_009749	5670433	94408188	14.19309822	13.62230775	0.570790468	1.485337179	-1.485337179	0.00010109	nucleus and cytoplasm, no further info
DCN	decorin	NM_007833	510332	145966821	9.445896272	10.09690164	-0.651005367	0.636836368	1.570262079	0.000150234	proteinaceous extracellular matrix, protein binding
EGR2	early growth response 2	NM_010118	3800403	118129947	8.379235968	7.99402626	0.385209708	1.306049619	-1.306049619	0.000152391	myelination, transcription regulation, metal ion binding
TNPO3	transportin 3	NM_177296	104610672	118130371	10.796607	11.21275929	-0.416152293	0.749420687	1.334364019	0.000175731	protein transport
SCL0001118.1_0 (Cbx3)	chromobox homolog 3 (Drosophila HP1 gamma)	AK002910.1	106940736	12833241	9.808515473	10.56602487	-0.757509396	0.591516618	1.69056958	0.000180104	Chromatin modification, Assembly and Disassembly
LOC674214	similar to 4933409K07Rik protein	XR_004625.1	1570138	94371793	9.89294188	9.272163805	0.620778075	1.537704274	-1.537704274	0.000192404	N/A
BOK	Bcl-2-related ovarian killer protein	NM_016778	1170373	118131164	11.66290462	10.94386979	0.699044836	1.625249063	-1.625249063	0.000197792	regulation and induction of apoptosis
SLC35F4	solute carrier family 35, member F4	NM_029238	106940180	142363008	9.550347081	8.942104299	0.608242782	1.52440134	-1.52440134	0.000204922	membrane transport
FBXO34	F-box protein 34	NM_030236	4010671	142357418	10.14893163	9.661429471	0.487502162	1.402015367	-1.402015367	0.000208089	ubiquitin cycle
MLLT4	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 4	NM_010806	106380402	145587091	9.74064496	10.15668572	-0.416040761	0.749478625	1.334260866	0.000251697	signal transduction, cell adhesion
MTAP1B	microtubule-associated protein 1B	NM_008634	2230332	6678945	10.41149898	11.01327888	-0.601779897	0.658940499	1.517587706	0.00035226	microtubule bundle formation, dendrite development, cytoskeletal regulatory protein binding
MLL5	myeloid/lymphoid or mixed-lineage leukemia 5	AK007682	104540338	51710848	9.672312414	10.2062917	-0.533979289	0.690647139	1.447917386	0.000392104	metal ion binding, regulation of transcription
PSMA4	proteasome (prosome, macropain) subunit, alpha type 4	NM_011966	4560427	118129984	11.81346406	12.31089916	-0.497435098	0.708365032	1.41170153	0.000393827	ubiquitin dependant protein catabolic process, threonine endopeptidase activity, hydrolase activity
EIF5B	eukaryotic translation initiation factor 5B	NM_198303	1190102	84043960	9.72288527	10.1021306	-0.379245327	0.768839665	1.300661302	0.000404827	translation initiation factor activity
DNAJB5	DnaJ (Hsp40) homolog, subfamily B, member 5	NM_019874	2900215	40254371	11.87937199	11.21824544	0.661126543	1.581316929	-1.581316929	0.000412731	heat shock protein binding
CORT	cortistatin	NM_007745	360008	142376224	9.256257056	8.664587004	0.591670051	1.506990217	-1.506990217	0.000469894	hormone activity, extracellular region
LOC100047645	similar to Dvl-binding protein Idax	XM_001478591.1	100430332	149251820	8.970502016	9.433201146	-0.46269913	0.725627415	1.378117721	0.00057552	N/A
LOC100046895	similar to Quaking protein	XM_001477026.1	105670403	149268446	9.353439578	9.962892966	-0.609453388	0.655444991	1.525681046	0.001008059	N/A
PEG3	paternally expressed 3	NM_008817	3710020	58331165	9.409115084	9.994403322	-0.585288238	0.666516161	1.500338715	0.001056082	apoptosis, metal ion binding
RDX	radixin	NM_001104616	4200039	6677698	9.016338627	9.472518783	-0.456180156	0.728913661	1.371904594	0.001096876	apical protein localisation, barbed end filament capping, microvillus biogenesis
6720463L11RIK (Sfrs7)	splicing factor, arginine/serine-rich 7	AK032861	105860154		9.099368447	9.582676125	-0.483307678	0.715335687	1.397945074	0.001202385	N/A
MIER3	mesoderm induction early response 1, family member 3	NM_172593	5290088	142388120	8.977309721	9.407669293	-0.430359572	0.74207681	1.347569399	0.001253947	regulation of transcription
EIF5	eukaryotic translation initiation factor 5	NM_173363	1230524	124430540	9.287482678	9.69405545	-0.406572772	0.754413411	1.325533169	0.001505808	GTP binding, translation initiation
RPP25	ribonuclease P 25 subunit (human)	NM_133982	6620048	19527203	10.2161968	9.79545962	0.420737176	1.338611372	-1.338611372	0.001526826	hydrolase activity, nucleic acid binding, tRNA processing
D130059O18RIK (Sfrs6)	splicing factor, arginine/serine-rich 6	AK051608	106100411		8.204036733	8.59406443	-0.390027697	0.763114954	1.310418561	0.001542882	N/A
EIF4EBP2	eukaryotic translation initiation factor 4E binding protein 2	NM_010124	5670279	118129916	9.236031806	9.635680707	-0.3996489	0.758042741	1.31918683	0.001568933	regulation of transcription initiation, insulin receptor signalling pathway, cAMP mediated signalling
JUNDM2	Ajuba	NP_034720.2	2360500	31982607	10.16148924	9.719410147	0.442079098	1.358560768	-1.358560768	0.00184923	regulation of transcription, protein binding
SSR2	signal sequence receptor, beta	NM_025448	5390180	141803157	11.06658225	11.56347776	-0.496895503	0.708630024	1.411173626	0.0020777	cotranslational protein targeting to membrane
LOC100041567	similar to p47 protein	XP_001476695.1	102260605	149258395	8.919220278	9.427182139	-0.50796186	0.70321519	1.422039817	0.002316301	N/A
CHMP6	chromatin modifying protein 6	NM_001085498	101770242	147904546	9.704337673	10.14869448	-0.444356803	0.734911889	1.360707337	0.002449243	protein transport, endosome
SERPINI1	serine (or cysteine) peptidase inhibitor, clade I, member 1	NM_009250	2940468	6678090	10.83727282	10.32312939	0.514143429	1.428145955	-1.428145955	0.002616102	serine type endopeptidase inhibitor activity
1110001P04RIK	RIKEN cDNA 1110001P04 gene	AK003265	103710563		9.921716777	10.32472887	-0.403012097	0.756277657	1.322265692	0.002966235	N/A
EGR4	early growth response 4	NM_020596	3120750	118130510	10.52968959	9.513770661	1.015918924	2.022190516	-2.022190516	0.003051492	DNA binding, metal ion binding, regulation of transcription-DNA dependant
RPS15A	ribosomal protein S15a	NM_170669	1170180	46358374	10.26062526	9.856671296	0.403953965	1.323129219	-1.323129219	0.003068986	structural constituent of ribosome, translation
GP1BB	glycoprotein Ib, beta polypeptide	NM_001001999	460725	50345975	10.4678908	9.902535315	0.565365485	1.47975208	-1.47975208	0.003219976	hemostasis, cell adhesion, protein binding
NIPBL	Nipped-B homolog (Drosophila)	NM_027707	5860592	49169844	9.36749506	9.768283335	-0.400788276	0.757444309	1.320229075	0.00361404	cell cycle, nucleus
MFN2	mitofusin 2	NM_133201	2260195	120407047	12.54701297	12.0845316	0.462481371	1.377909724	-1.377909724	0.003663364	multicellular organismal development, mitochondrial fusion, hydrolase activity, protein binding
SOSTDC1	sclerostin domain containing 1	NM_025312	6760017	142377757	8.851374589	9.330806767	-0.479432178	0.717259871	1.394194825	0.003720347	Wnt receptor signalling pathway, negative regulation of Bmp signalling pathway, protein binding
MSX1	homeobox, msh-like 1	NM_010835	2650309	113199782	8.651589718	9.037507712	-0.385917994	0.765291885	1.306690977	0.003841855	transcription, negative regulation of cell proliferation, negative regulation of apoptosis, mid/forebrain development, BMP signalling pathway, transcription repressor activity
SRRM1	serine/arginine repetitive matrix 1	NM_016799	6110025	142354391	10.02203481	10.47549238	-0.453457578	0.730290527	1.369318049	0.003889856	mRNA processing, RNA splicing
MTDH	Metadherin	NM_026002	70576	89363043	9.583120468	10.00942571	-0.426305244	0.744165161	1.343787713	0.003903988	regulation of transcription-DNA dependant, integral to membrane
SOCS2	suppressor of cytokine signaling 2	NM_007706	4760692	142379297	9.795042384	10.20316496	-0.408122575	0.753603424	1.326957877	0.004015711	positive regulation of neuron differentiation, negative regulation of signal transduction
CRELD1	cysteine-rich with EGF-like domains 1	NM_133930	2060253	19527147	10.69964816	10.18268611	0.516962048	1.430938877	-1.430938877	0.004098246	integral membrane protein, calcium ion binding
WBP4	WW domain binding protein 4	NM_018765	2230019	124286863	9.836538753	10.27399483	-0.437456076	0.738435555	1.354214316	0.004629468	mRNA processing and RNA splicing
ARC	activity regulated cytoskeletal-associated protein	NM_018790	4610093	86604725	11.0425753	9.971392365	1.071182939	2.101155505	-2.101155505	0.004654732	endocytosis, multicellular organismal development
PSMB7	proteasome (prosome, macropain) subunit, beta type 7	NM_011187	5570088	6755205	11.85188468	12.34261916	-0.490734485	0.711662694	1.405160069	0.005119333	ubiquitin dependant protein catabolic process, threonine endopeptidase activity, hydrolase activity

Table 4.3: Complete list of dysregulated genes in the L100P adult group. Cont next page…

TargetID	Full Gene Name	Accession Number	ProbeID	GI numbers	C57Sal MEAN expression	L100PSal MEAN expression	Difference	anti-log	fold change	Pval	GO Terms
TRAK2	trafficking protein, kinesin binding 2	NM_172406	430736	40254202	10.60403212	10.16698748	0.437044644	1.353828172	-1.353828172	0.005179239	GABA receptor binding
EGR3	early growth response 3	NM_018781	6940128	118130757	11.52153466	10.51817594	1.003358718	2.004661596	-2.004661596	0.005446415	transcription, PNS development, neuromuscular synaptic transmission, metal ion binding
MED4	mediator of RNA polymerase II transcription, subunit 4 homolog (yeast)	NM_026119	4480050	118129849	9.856894195	10.25107521	-0.394181015	0.760921212	1.314196509	0.005495297	transcription, mediator complex, protein binding
UBE1X	ubiquitin-like modifier activating enzyme 1	NM_009457.3	4150139	141803326	10.35694711	10.89095819	-0.534011077	0.690631922	1.447949289	0.006125225	protein modification, ubiquitin cycle, ligase activity
1200009O22RIK	RIKEN cDNA 1200009O22 gene	NM_025817	4780242	31541813	10.94152709	11.37844369	-0.436916591	0.738711739	1.353708013	0.006451177	N/A
HIST1H2BC	histone cluster 1, H2bc	NM_023422	2570156	142381596	10.9595606	10.26174085	0.697819748	1.622051637	-1.622051637	0.006533336	nucleosome assembly, DNA binding
POU3F1	POU domain, class 3, transcription factor 1	NM_011141	3710022	145279231	11.02723439	10.59858935	0.428645039	1.345968866	-1.345968866	0.006759471	DNA dependant transcription regulation, myelination, keratinocyte differentiation, transcription factor activity, DNA binding
AMMECR1L	AMME chromosomal region gene 1-like	NM_153515	1780270	118130140	9.446718999	9.880277517	-0.433558518	0.740433196	1.350560733	0.006855014	N/A
SCL0004190.1_3 (MII5)	myeloid/lymphoid or mixed-lineage leukemia 5	AK007682.1	100050026	12841381	9.953044302	10.45541926	-0.502374955	0.705943704	1.416543549	0.006934314	N/A
VASN	vasorin	NM_139307	1770722	31981692	9.157010126	9.593924862	-0.436914735	0.738712689	1.353706271	0.007019803	protein binding
RGL1	ral guanine nucleotide dissociation stimulator,-like 1	NM_016846	1500097	118131206	10.42226476	10.03822491	0.384039853	1.304990996	-1.304990996	0.007346557	regulation of small GTPase mediated signal transduction, guanyl-nucleotide exchange factor activity
AI854703	expressed sequence AI854703	AK038960	4810739	149255141	10.19317863	9.768674694	0.424503941	1.342110949	-1.342110949	0.00787173	N/A
SYT16	synaptotagmin XVI	NM_172804	2690504	121674808	11.37073419	10.90823686	0.462497333	1.37792497	-1.37792497	0.007965758	protein binding, phospholipid binding, protein homo/heterodimerisation activity
CD63	Cd63 antigen	NM_001042580	6660451	110431343	11.41740475	11.94329399	-0.525889247	0.694530884	1.439820782	0.008217723	late endosome, lysosome, integral to membrane
WDR6	WD repeat domain 6	NM_031392	7050154	31981423	10.39012413	10.7858959	-0.39577177	0.760082662	1.315646377	0.008256377	Contains WD40/YVTN repeat-like and WD40 repeat, no other info available
HSP110	heat shock 105kDa/110kDa protein 1	NP_038587.2	1690017	114145504	13.53428075	12.87490485	0.659375896	1.579399235	-1.579399235	0.008285994	response to stress, response to heat, ATP binding
CSNK1E	casein kinase 1, epsilon	NM_013767	2850347	146134466	10.20815214	10.69135902	-0.483206872	0.715386571	1.397847399	0.008833556	Wnt receptor signalling pathway, protein amino acid phosphorylation, circadian rhythm, transferase activity, kinase activity, ATP binding
EZH2	enhancer of zeste homolog 2 (Drosophila)	NM_007971	6130605	6679720	8.033558632	8.55138207	-0.517823438	0.698424737	1.431793502	0.008854039	transcription, negative regulation of striated muscle cell differentiation, histone methylation
AI448196 (Armcx4)	armadillo repeat containing, X-linked 4	AI448196	106380397	38087316	9.876833887	10.32277155	-0.445937663	0.734107037	1.362199175	0.009021746	N/A
APRIN (Pds5)	PDS5, regulator of cohesion maintenance, homolog B (S. cerevisiae)	NM_175310	2810050	118129869	9.153794491	9.590320487	-0.436525996	0.738911765	1.35334156	0.009097201	cell division, regulation of cell proliferation
VILLP	villin-like	NM_011700	130672	23956063	8.294780283	8.730621053	-0.435840771	0.739262803	1.352698926	0.009258746	cytoskeletal organisation and biogenesis
CRTC1	CREB regulated transcription coactivator 1	NM_001004062	2940097	118131085	9.778775522	9.391595665	0.387179857	1.307834383	-1.307834383	0.009271174	DNA dependant transcription regulation
FDPS	farnesyl diphosphate synthetase	NM_134469	60022	118129959	10.531709	10.91393794	-0.382228939	0.767251284	1.30335396	0.009291575	steroid biosynthetic process, sterol biosynthetic process, isoprenoid biosynthetic process, lipid biosynthetic process, transferase activity
SCL0002702.1_3805		AK083781.1	103130204	26101487	8.91404176	9.321351458	-0.407309698	0.754028157	1.326210422	0.009339525	N/A
NPTX2	neuronal pentraxin 2	NM_016789	70021	31560467	9.459919151	8.991576224	0.468342926	1.38351945	-1.38351945	0.009800026	calcium ion binding, metal ion binding, sugar binding, extracellular
NPY	neuropeptide Y	NM_023456	3170138	27754168	12.41875305	11.54356996	0.875183094	1.834240857	-1.834240857	0.009992619	neuropeptide signalling pathway, regulation of blood pressure, G-protein coupled receptor binding, extracellular
REC8	REC8 homolog (yeast)	NM_020002	5900014	31982698	8.566953493	8.174879246	0.392074246	1.312278787	-1.312278787	0.010025929	chromosome segregation, meiosis
LOC100046770	similar to histone macroH2A1.2	XM_001476780.1	101850148	149264153	9.898160405	10.38452257	-0.486362165	0.713822774	1.400907952	0.010123106	N/A
RYK	receptor-like tyrosine kinase	NM_001042607	1500338	110681701	10.07179275	10.57529814	-0.503505392	0.705390772	1.41765393	0.010332659	axonogenesis, protein amino acid phosphorylation, Wnt receptor activity
SLIT1	slit homolog 1 (Drosophila)	NM_015748	50110	141801919	9.311456536	8.923616123	0.387840412	1.308433328	-1.308433328	0.010528995	axon guidance, cell differentiation, nervous system development, nucleus localization
GTPBP4	GTP binding protein 4	NM_027000	6840332	146141210	9.822670827	10.24973485	-0.427064021	0.743773875	1.344494656	0.010563211	ribosome biogenesis and assembly, translation
MAST3	microtubule associated serine/threonine kinase 3	AK140821	450402	40656287	14.07360221	13.0793661	0.994236109	1.992025491	-1.992025491	0.010920477	ATP binding, protein amino acid phosphorylation, transferase activity, protein serine/threonine kinase activity, nucleotide binding, magnesium ion binding
GJB6	gap junction protein, beta 6	NM_001010937	5390441	58331125	12.29594005	11.48516568	0.810783472	1.754163802	-1.754163802	0.010927777	sensory perception of sound, cell-cell signalling, integral to membrane, gap junction
PIP5K2A	phosphatidylinositol-4-phosphate 5-kinase type II alpha	NC_007866.1	5220497	126012549	12.37932134	11.82891064	0.550410697	1.464502542	-1.464502542	0.011047484	phosphatidylinositol metabolic process, transferase activity, kinase activity
LOC675440	similar to SMT3 suppressor of mif two 3 homolog 2 (S. cerevisiae)	XM_001479274.1	70010	149249298	11.12408024	11.54225254	-0.4181723	0.74837211	1.336233654	0.011211119	Regulation of transcription, DNA repair, protein modification
NPM3-PS1	nucleoplasmin 3, pseudogene 1	NR_002702	3610300	84794553	9.007533266	9.485276206	-0.47774494	0.7180992	1.392565262	0.011533683	N/A
CADPS	Ca2+-dependent secretion activator	NM_001042617	2320181	110825965	12.88094852	12.18282592	0.698122599	1.622392175	-1.622392175	0.011909626	calcium ion binding, synapse, protein transport, vesicle organisation and biogenesis, catecholamine secretion
D4BWG0951E	DNA segment, Chr 4, Brigham & Women's Genetics 0951 expressed	NM_026821	4570440	146149309	10.5435239	10.0128078	0.530716107	1.444646093	-1.444646093	0.011992251	N/A
MFGE8	milk fat globule-EGF factor 8 protein	NM_001045489	3440373	113865978	12.06467704	11.50905378	0.555623259	1.469803467	-1.469803467	0.012169796	positive regulation of phagocytosis, cell adhesion, external side of plasma membrane
CCND2	cyclin D2	NM_009829	5340167	80751174	10.03479455	10.7427288	-0.707934256	0.612196093	1.633463544	0.012192576	cell division, regulation of cell cycle
KIF1A	kinesin family member 1A	NM_001110315	450632	141801924	9.785736454	9.380839347	0.404897108	1.32399448	-1.32399448	0.012217286	ATP binding, protein transport, microtubule processes
PLEKHA2	pleckstrin homology domain-containing, family A (phosphoinositide binding specific) member 2	NM_031257	2850091	31543492	10.8350422	10.26304013	0.572002076	1.486585124	-1.486585124	0.012220196	lipid binding, phosphatidylinositol binding
GALK1	galactokinase 1	NM_016905	840162	93102412	9.041578843	9.506174677	-0.464595834	0.724674064	1.379930716	0.012390252	phosphporylation, metabolic process, transferase activity, protein binding, kinase activity, nucleotide binding
REEP3	receptor accessory protein 3	NM_178606	3190239	146198669	10.02322305	10.44283764	-0.419614592	0.747624321	1.337570183	0.012433073	integral membrane protein
SELPLG	selectin, platelet (p-selectin) ligand	NM_009151	1770167	31982018	10.40434578	9.744762059	0.659583724	1.579626772	-1.579626772	0.012448906	leukocyte tethering or rolling, protein binding, sugar binding, cell adhesion
PCBP1	poly(rC) binding protein 1	NM_011865	1050088	47271536	10.46528938	10.98542711	-0.520137724	0.697305263	1.434092144	0.012651694	mRNA processing, translation activator activity, ribonucleoprotein complex
TPM3	tropomyosin 3, gamma	NM_022314	670600	40254524	9.18351663	9.569745866	-0.386229236	0.765126801	1.306972908	0.012692399	actin binding, regulation of muscle contraction
BXDC2	brix domain containing 2	NM_026396	5550215	141802629	9.274177688	9.694594541	-0.420416853	0.747208694	1.338314192	0.012740641	ribosome biogenesis and assembly, translation
WSB1	WD repeat and SOCS box-containing 1	NM_001042565	870563	141803466	9.628439911	10.16947351	-0.541033602	0.68727834	1.455014572	0.012891541	intracellular signalling cascade
RPS27L	ribosomal protein S27-like	NM_026467	50577	141803225	11.02492363	11.68446175	-0.659538115	0.633080948	1.579576835	0.012977736	structural constituent of ribosome, translation, metal ion binding, ribonucleoprotein complex
CECR6	cat eye syndrome chromosome region, candidate 6 homolog (human)	NM_033567	540672	15808997	10.28384918	9.834651315	0.449197863	1.36528095	-1.36528095	0.013031014	regulation of transcription, integral to membrane
BTBD6	BTB (POZ) domain containing 6	NM_201646	110088	42517137	11.03804991	10.63367339	0.404376518	1.323516809	-1.323516809	0.013031719	protein binding
ASNS	asparagine synthetase	NM_012055	110368	146134364	10.1676237	9.779771847	0.387851868	1.308443708	-1.308443708	0.013037244	metabolic processes, amino acid biosynthetic process
ZFP292	zinc finger protein 292	NM_013889	840736	149252215	9.197168665	9.606497399	-0.409328734	0.75297364	1.328067739	0.013259523	nucleic acid binding
ATP1B2	ATPase, Na+/K+ transporting, beta 2 polypeptide	NM_013415	4480603	146134451	12.07376971	11.13988344	0.933886267	1.910415261	-1.910415261	0.013920062	ion transport
NOL5A	nucleolar protein 5A	AK009799	5570139	126090931	9.167593006	9.561547258	-0.393954253	0.761040822	1.31398996	0.014094517	ribosome biogenesis and assembly
LARGE	like-glycosyltransferase	NM_010687	6940068	6754505	11.25540446	10.86880924	0.386595219	1.307304503	-1.307304503	0.014321744	transferase activity and biosynthetic processes
GSTM5	glutathione S-transferase, mu 5	NM_010360	6550446	133892313	12.79985355	12.40087686	0.398976685	1.318572305	-1.318572305	0.014386257	metabolic process, glutathione transferase activity
GPSM1	G-protein signalling modulator 1 (AGS3-like, C. elegans)	NM_153410	2230735	146149078	9.673586944	10.10050645	-0.426919502	0.743848385	1.344359981	0.014461886	signal transduction, NS development, cell differentiation, GTPase activator activity

Table 4.3: Complete list of dysregulated genes in the L100P adult group. Continued next page….

TargetID	Full Gene Name	Accession Number	ProbeID	GI numbers	C57Sal MEAN expression	L100PSal MEAN expression	Difference	anti-log	fold change	Pval	GO Terms
NME2	non-metastatic cells 2, protein (NM23B) expressed in	NM_001077529	2190672	117606273	12.31092296	12.82784863	-0.516925675	0.698659489	1.4309028	0.01464483	UTP biosynthetic process, nucleotide metabolic process, GTP/CTP biosynthetic process, transferase activity, kinase activity, protein binding
RALYL	RALY RNA binding protein-like	NM_178631	4210471	142362122	11.7512141	11.13902366	0.612190436	1.52857828	-1.52857828	0.01464896	nucleic acid binding
BCAS1	breast carcinoma amplified sequence 1	NM_029815	2030711	30023813	12.19098224	11.10330821	1.087674023	2.12531108	-2.12531108	0.014884933	protein homodimerization
SORCS3	sortilin-related VPS10 domain containing receptor 3	NM_025696	5700309	118130290	10.22682518	9.792214245	0.43461093	1.351546295	-1.351546295	0.014984355	membrane
LOC385905	similar to proteasome alpha7/C8 subunit	XR_034995.1	100540242	149267541	10.9690955	11.40714364	-0.438048137	0.738132574	1.35477018	0.015035484	N/A
PSMA1	proteasome (prosome, macropain) subunit, alpha type 1	NM_011965	380059	33563281	9.641608393	10.14876072	-0.507152328	0.703609893	1.421242098	0.01508205	ubiquitin dependant protein catabolic process, threonine endopeptidase activity, hydrolase activity
HAPLN4	hyaluronan and proteoglycan link protein 4	NM_177900	4120112	118130529	10.33542115	9.588874902	0.746546246	1.677771501	-1.677771501	0.015187263	hyaluronic acid binding, cell adhesion, extracellular
LOC100044163	hypothetical protein LOC100044163	XP_001471724.1	3140131	149263016	12.80463309	11.99934153	0.805291558	1.747498913	-1.747498913	0.015380651	N/A
GAD1	glutamic acid decarboxylase 1	NM_008077	2360035	145301579	11.179378	10.49138948	0.687988522	1.611035765	-1.611035765	0.01542056	synaptic transmission, neurotransmitter biosynthetic process, carboxylic acid metabolic process, pyridoxal phosphate binding
ALDH1A1	aldehyde dehydrogenase family 1, subfamily A1	NM_013467	6520706	118129922	9.495910579	8.908435481	0.587475097	1.502614675	-1.502614675	0.015437916	oxidoreductase activity, metabolic process
PRKCB1	protein kinase C, beta 1	NM_008855	3130092	116734871	13.24611142	12.10277157	1.143339852	2.208917983	-2.208917983	0.015450857	protein amino acid phosphorylation, cellular calcium ion homeostasis, protein kinase activity, calcium channel regulator activity, metal ion binding, protein binding
CUEDC2	CUE domain containing 2	NM_024192	460095	13195619	10.91137463	10.50795	0.403424631	1.322643844	-1.322643844	0.015492211	ubiquitin cycle
HES6	hairy and enhancer of split 6 (Drosophila)	NM_019479	540411	142369243	9.991146841	10.72797986	-0.736833017	0.60005514	1.666513514	0.015531095	DNA dependant regulation of transcription, NS development, cell differentiation
COL8A2	collagen, type VIII, alpha 2	NM_199473	430020	41054949	8.08808554	8.620286029	-0.534200489	0.690541254	1.448139404	0.015543422	phosphate transport, epithelial cell proliferation, cell adhesion, camera-type eye morphogenesis, protein binding
AW551984	expressed sequence AW551984	NM_178737	2350563	146198873	7.924408601	8.330647398	-0.406238797	0.754588073	1.325226352	0.015864086	integrin-mediated signaling pathway
SLC25A23	solute carrier family 25 (mitochondrial carrier, phosphate carrier), member 23	NM_026877	3440441	70980533	9.859672348	9.406769193	0.452903156	1.368791926	-1.368791926	0.015910869	mitochondrial inner membrane, calcium ion binding, transporter activity
CHD7	chromodomain helicase DNA binding protein 7	NM_001081417	3870372	124487248	9.145392262	9.744940457	-0.599548195	0.659960601	1.515241968	0.015975237	transcription, sensory perception of sound, locomotor behaviour, chromatin modification, DNA binding, helicase activity, nucleotide binding
TBCA	tubulin cofactor a	NM_009321	6400095	118130449	12.33848967	12.79720529	-0.458715626	0.727633754	1.374317773	0.016169517	tubulin folding
SLC2A6	solute carrier family 2 (facilitated glucose transporter), member 6	NM_172659	450609	153945871	10.48970598	9.810103196	0.67960278	1.601698696	-1.601698696	0.016309302	carbohydrate transport
DDB1	damage specific DNA binding protein 1	NM_015735	6110687	7657010	11.33381441	11.80440593	-0.470591528	0.721668641	1.385677502	0.016585716	ubiquitin cycle, nucleic acid binding
CORO1A	coronin, actin binding protein 1A	NM_009898	3140609	31982807	11.0419502	10.58201013	0.459940065	1.375484674	-1.375484674	0.016746202	uropod organisation and biogenesis, T cell homeostasis, regulation of actin polymerisation and/or depolymerisation, leukocyte chemotaxis, colocalises with plasma membrane, actin binding
AKAP9	A kinase (PRKA) anchor protein (yotiao) 9	NM_194462	510184	125661047	9.595622179	9.976576138	-0.380953959	0.767929641	1.302202632	0.016816231	kinase activity, protein binding
RASD1	RAS, dexamethasone-induced 1	NM_009026	2450463	145966901	9.238041161	8.778763684	0.459277477	1.374853099	-1.374853099	0.016954596	small GTPase mediated signal transduction, protein transport
RGS10	regulator of G-protein signalling 10	NM_026418	2340292	142388581	9.821631642	9.285046163	0.536585479	1.450535374	-1.450535374	0.017028404	negative regulation of signal transduction, protein binding
ITPKA	inositol 1,4,5-trisphosphate 3-kinase A	NM_146125	6220075	118130226	12.76015278	11.4013748	1.358777979	2.564678488	-2.564678488	0.017094529	inositol metabolic process, nucleotide binding, calmodulin binding, transferase activity
SLC6A17	solute carrier family 6 (neurotransmitter transporter), member 17	NM_172271	6200138	142370279	11.48192521	10.71465031	0.767274897	1.702051743	-1.702051743	0.017133097	neurotransmitter transport
1500010J02RIK	RIKEN cDNA 1500010J02 gene	NM_001013256	2630138	61742799	10.36963803	10.81034752	-0.440709494	0.736772187	1.357271647	0.017294124	integral membrane protein
HIRIP3	HIRA interacting protein 3	NM_172746	940066	34326383	9.369847251	9.844650809	-0.474803558	0.719564762	1.389726977	0.01734253	N/A
MTDNA_ND2			103290091		14.33955411	13.61089996	0.728654154	1.657092519	-1.657092519	0.01750446	N/A
3110018K01RIK (Sept3)	Septin 3	NM_011889	106590102		12.83081861	12.34212713	0.488691477	1.403171623	-1.403171623	0.017623295	Cell cycle, cell division, septin cycle
TTYH1	tweety homolog 1 (Drosophila)	NM_001001454	3780672	141803438	11.24767778	10.82292053	0.424757246	1.342346615	-1.342346615	0.017646958	cell adhesion, ion transport
SLC25A12	solute carrier family 25 (mitochondrial carrier, Aralar), member 12	NM_172436	7400731	141802710	10.66823247	10.19165064	0.476581834	1.391443023	-1.391443023	0.0176937	malate-aspartate shuttle, metal ion binding
USP20	ubiquitin specific peptidase 20	NM_028846	520128	142359890	11.8277114	11.35667612	0.471035281	1.386103782	-1.386103782	0.017897436	metal ion binding, pepsidase and hydrolase activity, ubiquitin cycle and ubiquitin-dependant catabolic processes
CLDN11	claudin 11	NM_008770	670338	118130649	14.3809118	13.74119833	0.639713472	1.558019696	-1.558019696	0.017963611	cell adhesion, axon ensheathment, protein binding, integral to membrane
DCT	dopachrome tautomerase	NM_010024	1090347	142352300	7.881760935	8.311225723	-0.429464788	0.742537201	1.346733872	0.017989092	metabolic process, melanin biosynthetic process, cell development, metal ion binding, isomerase activity
TDE2 (serinc1)	serine incorporator 1	NC_000076.5	6650020	9790268	14.02636788	13.19829698	0.8280709	1.775309918	-1.775309918	0.018109743	Lipid metabolism
ADCY2	adenylate cyclase 2	NM_153534	4610100	124244091	12.12471967	11.33975589	0.78496378	1.723049061	-1.723049061	0.01815896	G-protein signaling, coupled to cAMP nucleotide second messenger
NUP85	nucleoporin 85	NM_001002929	105570044	108773812	10.4426203	10.88637901	-0.443758703	0.735216626	1.360143343	0.018227095	protein transport, macrophage chemotaxis, lamellipodium biogenesis, cytokine and chemokine mediated signalling pathway, nuclear pore
PLEKHB1	pleckstrin homology domain containing, family B (evectins) member 1	NM_013746	540546	7305378	11.85449211	10.79972548	1.05476663	2.077382142	-2.077382142	0.018420383	phosphorecepter outer segment, protein binding
TMEM46	shisa homolog 2	NM_145463.4	450672	141803556	8.584316904	9.121191839	-0.536874935	0.689262325	1.450826433	0.018478114	integral membrane protein, multicellular organismal development
NOL5	nucleolar protein 5	NM_018868	580731	120407049	9.54685735	10.00466921	-0.457811864	0.728089716	1.373457115	0.018649577	ribosome biogenesis and assembly, nucleic acid binding
FSTL1	follistatin-like 1	NM_008047	2900113	145966858	8.796681707	9.59374448	-0.797062773	0.575519704	1.737559971	0.018899831	calcium ion binding, heparin binding, extracellular
BC055107	cDNA sequence BC055107	NM_183187	6180088	118130726	13.05605993	11.79599107	1.26006886	2.395071724	-2.395071724	0.019107883	N/A
A730063M14RIK	RIKEN cDNA A730063M14 gene	AK080508	102360129		9.138245945	9.522222682	-0.383976737	0.766322337	1.304933906	0.019198989	N/A
H2AFY	H2A histone family, member Y	NM_012015	4730750	41152516	10.87385441	11.34791988	-0.47406547	0.719932988	1.389018169	0.019214656	chromosome organisation and biogenesis, dosage compensation
GPR68	G protein-coupled receptor 68	NM_175493	1190537	31342018	9.133661897	8.753007932	0.380653965	1.30193188	-1.30193188	0.019226438	g-protein coupled receptor activity, signal transduction, inflammatory response
GNB1	guanine nucleotide binding protein (G protein), beta 1	NM_008142	2120397	111186467	10.53468166	11.09286141	-0.55817975	0.679158518	1.472410303	0.019239266	sensory perception of taste, Ras protein signal transduction, G-protein coupled receptor signalling pathway, muscarinic pathway, cell proliferation
DLG2	discs, large homolog 2 (Drosophila)	NM_011807	4230095	118136296	10.76755712	10.13413035	0.633426765	1.55124522	-1.55124522	0.019313055	protein binding, sensory perception of pain, synaptic transmission
LOC100044324	similar to euchromatic histone methyltransferase 1	XM_001471971.1	5910086	149249333	9.070392315	9.497159226	-0.426766911	0.743927064	1.344217798	0.019349865	N/A
S100A1	S100 calcium binding protein A1	NM_011309	6770707	113930760	11.39847073	10.54569908	0.852771645	1.805967143	-1.805967143	0.019430186	metal ion binding, calcium ion binding
WNT5A	wingless-related MMTV integration site 5A	NM_009524	840685	46909566	8.438232177	8.830799656	-0.392567478	0.761772716	1.312727509	0.01947665	cell signalling, neurotransmitter secretion, signal transduction, synapse organisation, embryonic development
PPP1R9A	protein phosphatase 1, regulatory (inhibitor) subunit 9A	NM_181595	5390408	118130685	12.76627393	12.21936144	0.546912494	1.460955762	-1.460955762	0.019525315	actin filament organisation, neurite development, cytoskeleton, protein binding
AU040829	expressed sequence AU040829	NM_001099288	3830731	151108448	12.78324145	11.90945401	0.873787435	1.832467276	-1.832467276	0.019548952	signal transduction
LOC268569	hypothetical gene supported by NM_016750	XM_195821.2	103840092	38050556	11.42330866	12.24310458	-0.819795917	0.568522077	1.765156277	0.019582555	N/A
KCNQ2	potassium voltage-gated channel, subfamily Q, member 2	NM_001003824	130204	54873653	10.25856539	9.767906944	0.490658449	1.405086014	-1.405086014	0.019643804	cation transport, voltage-gated potassium channel complex
SLC8A2	solute carrier family 8 (sodium/calcium exchanger), member 2	NM_148946	6220092	40254192	9.42508631	9.012858283	0.412228026	1.330739356	-1.330739356	0.019744173	transmembrane calcium ion transport
2810410M20RIK (Dpy30)	dpy-30 homolog (C. elegans)	NM_024428	2510025	142348817	10.0685203	10.46488601	-0.396365709	0.75976981	1.316188124	0.019827415	nucleus
SCL0003457.1_572 (mobp)	myelin-associated oligodendrocytic basic protein	AK083103.1	100840184	26350242	12.41453659	11.09788474	1.316653849	2.490877114	-2.490877114	0.020086006	structural constituent of myelin sheath
9330159F19RIK	RIKEN cDNA 9330159F19 gene	AK034141	2100739	26495280	8.664058285	8.271504966	0.392553319	1.312714625	-1.312714625	0.020095755	N/A

Table 4.3: Complete list of dysregulated genes in the L100P adult group. Cont next page…

TargetID	Full Gene Name	Accession Number	ProbeID	GI numbers	C57Sal MEAN expression	L100PSal MEAN expression	Difference	anti-log	fold change	Pval	GO Terms
ANK	progressive ankylosis	NM_020332	6400465	142366627	12.3750122	11.64751655	0.727496666	1.655762392	-1.655762392	0.020106959	phosphate transport, regulation of bone mineralization
GABRA1	gamma-aminobutyric acid (GABA-A) receptor, subunit alpha 1	NM_010250	130372	145966746	10.68090111	9.764415572	0.916485541	1.88751164	-1.88751164	0.02016683	Synaptic transmission, ion transport, neurotransmitter receptor activity
POLR2D	polymerase (RNA) II (DNA directed) polypeptide D	NM_027002	130280	146141112	9.340948751	9.78296631	-0.44201756	0.736104472	1.358502819	0.020256	transcription, DNA-directed RNA polymerase action
PRR18	proline rich region 18	NM_178774	2030064	40254340	10.39053255	9.981408397	0.409124158	1.327879431	-1.327879431	0.020273168	endoplasmic reticulum
TUBB5	tubulin, beta 5	NM_011655	730670	142374674	12.59679691	13.01888318	-0.422086275	0.746344558	1.339863725	0.020290257	microtubule based movement, protein polymerisation
HSPA2	heat shock protein 2	NM_001002012	6420592	50345977	12.79027848	11.86596542	0.924313058	1.897780395	-1.897780395	0.020334836	ATP binding, protein binding, heat shock protein
ST8SIA5	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 5	NM_153124	3520333	24233503	9.313640651	8.868600597	0.445040054	1.361351912	-1.361351912	0.020402216	protein amino acid glycosylation, transferase activity
RHBDF1	rhomboid family 1 (Drosophila)	NM_010117	380538	32306546	9.368584595	9.766038956	-0.397454361	0.759196706	1.317181689	0.020546359	receptor activity
RDBP	RD RNA-binding protein	NM_001045863	6590044	114052413	10.13390611	10.52978448	-0.396878372	0.7600265	1.315743595	0.020619571	regulation of transcription, nucleotide binding
SNPH	syntaphilin	NM_198214	2680484	118130824	13.58195386	12.26781776	1.3141361	2.4865339	-2.4865339	0.020646038	protein binding, synaptosome, vessicle fusion, negative regulation of endocytosis
BZW2	basic leucine zipper and W2 domains 2	NM_025840	940079	31981159	10.57690947	11.16342217	-0.586512702	0.665950706	1.501612643	0.02083823	regulation of translational initiation, NS development, cell differentiation
GSTM1	glutathione S-transferase, mu 1	NM_010358	1940332	141803498	11.03462754	10.50990727	0.524720277	1.438654613	-1.438654613	0.021084051	metabolic process, transferase activity
FKBP9	FK506 binding protein 9	NM_012056	1400044	27532951	9.86754793	10.48275135	-0.615203423	0.652837836	1.531773965	0.021150623	Calcium ion binding, isomerase activity, protein folding
ALDOC	aldolase 3, C isoform	NM_009657	450121	38091506	13.0887938	12.06341149	1.025382312	2.035498717	-2.035498717	0.021154484	glycolysis, metabolic process
TUBB4	tubulin, beta 4	NM_009451	5310681	31981938	12.72910575	11.90653612	0.822569627	1.768553212	-1.768553212	0.021339599	GTP binding, GTPase activity, nucleotide binding, microtubule-based processes and protien polymerisation
HNRPH1	heterogeneous nuclear ribonucleoprotein H1	NM_021510	1170086	125628660	9.540625977	10.22682613	-0.686200151	0.621488607	1.609039955	0.021641304	RNA splicing, mRNA processing, nucleotide binding
0610010I05RIK (mt-Nd5)	NADH dehydrogenase 5, mitochondrial	AK137271	100450066		13.53991229	12.94198198	0.597930309	1.513543678	-1.513543678	0.021680037	ATP synthesis coupled electron transport, mitochondrial
RYR2	ryanodine receptor 2, cardiac	NM_023868	2760671	124430577	9.810305524	9.218438471	0.591867054	1.507196013	-1.507196013	0.021735728	ion transport, response to caffine, protein binding, integral to membrane
2310036D22RIK (Tmem106b)	transmembrane protein 106b	NM_027992	104810195	21624614	11.06351572	10.51973997	0.543775751	1.457782767	-1.457782767	0.021760596	Integral membrane protein
NRGN	neurogranin	NM_022029	4280433	145587078	13.42841287	11.93096359	1.497449276	2.823430807	-2.823430807	0.021773197	protein kinase cascade, calmodulin binding
ORC6L	origin recognition complex, subunit 6-like (S. cerevisiae)	NM_019716	2260592	9790136	9.948640571	10.38709236	-0.438451793	0.737926078	1.355149288	0.021806166	DNA replication, DNA binding, nuclear origin of replication recognition complex
IDH1	isocitrate dehydrogenase 1 (NADP+), soluble	NM_001111320	1990021	58037545	10.24147769	10.82001193	-0.578534237	0.669643783	1.493331269	0.021927295	tricarboxylic acid cycle, response to oxidative stress, manganese ion binding, metal ion binding, isocitrate dehydrogenase (NADP+) activity
ANKRD34	ankyrin repeat domain 34A	NM_001024851	5910176	122114647	11.32170023	10.63215856	0.689541673	1.612771079	-1.612771079	0.021935617	transcription regulation
SCD1	stearoyl-Coenzyme A desaturase 1	NM_009127	2680441	118130513	11.46687962	10.93189328	0.534966346	1.448928442	-1.448928442	0.02194752	lipid biosynthesis, lipid metabolism, stearoyl-CoA 9-desaturase activity
RNASEN	ribonuclease III, nuclear	NM_026799	105720279	54873614	9.676635221	10.05665742	-0.3800222	0.768425766	1.30136188	0.02199252	ribonuclease III activity
ATP2A2	ATPase, Ca++ transporting, cardiac muscle, slow twitch 2	NM_001110140	1090075	142386634	13.49624141	12.68659902	0.809642391	1.752776918	-1.752776918	0.022210058	cellular calcium ion homeostasis, calcium ion transport
ID1	inhibitor of DNA binding 1	NM_010495	360398	118130032	9.016825115	9.40114191	-0.384316795	0.766141728	1.305241528	0.022284642	regulation of transcription, regulation of angiogenesis, multicellular organismal development
SPINK8	serine peptidase inhibitor, Kazal type 8	NM_183136	4670086	114145595	10.14366752	9.434450077	0.709217444	1.634917054	-1.634917054	0.022370146	endopeptidase inhibitor
HCN2	hyperpolarization-activated, cyclic nucleotide-gated K+ 2	NM_008226	3840181	88014633	11.3542375	10.46339573	0.890841768	1.85425771	-1.85425771	0.022590024	regulation of membrane potential, sodium ion transport, potassium ion transport, cAMP binding
EDG8 (S1pr5)	sphingosine-1-phosphate receptor 5	NM_053190	1990441	16716488	9.360120593	8.981289622	0.37883097	1.300287793	-1.300287793	0.022610735	signal transduction, G-protein coupled receptor protein signalling pathway
TGFBI	transforming growth factor, beta induced	NM_009369	2060446	145966862	8.312869316	8.939348098	-0.626478782	0.647755479	1.543792423	0.022706007	protein binding, cell adhesion, extracellular
SEC14L2	SEC14-like 2 (S. cerevisiae)	NM_144520	2640370	118130014	9.452180209	9.018470333	0.433709875	1.350702432	-1.350702432	0.022818649	lipid binding, regulation of transcription, transport
AQP1	aquaporin 1	NM_007472	450338	145966713	8.548667263	8.978599238	-0.430031975	0.742245335	1.347263436	0.022885042	water channel activity
2010300C02RIK	RIKEN cDNA 2010300C02 gene	AK008485	5690551	149233727	12.08794583	10.99080387	1.097141955	2.139304663	-2.139304663	0.022898871	N/A
CAMK2A	calcium/calmodulin-dependent protein kinase II alpha	NM_009792	1740333	118130259	12.52835568	11.28941935	1.238936328	2.360244517	-2.360244517	0.022981063	regulation of neurotransmitter secretion, protein amino acid phosphorylation, calcium ion transport, calcium and calmodulin dependant protein kinase complex, transferase activity, synapse
BCL6	B-cell leukemia/lymphoma 6	NM_009744	940100	142360700	11.47595964	10.73979515	0.736164489	1.665741451	-1.665741451	0.023110102	anti-apoptosis, axon regeneration
PDXP	pyridoxal (pyridoxine, vitamin B6) phosphatase	NM_020271	5050204	85838507	12.90401128	12.25514888	0.648862408	1.567931365	-1.567931365	0.023121659	metabolic process, catalytic activity, metal ion binding, phosphoric monoester hydrolase activity
UBL3	ubiquitin-like 3	NM_011908	6450458	134053909	12.89170182	12.34810167	0.543600142	1.457605333	-1.457605333	0.023371204	protein modification
MYH8	myosin, heavy polypeptide 8, skeletal muscle, perinatal	NM_177369	100450446	71143151	8.397948894	9.181007131	-0.783058237	0.581133595	1.720774721	0.023655088	Actin binding, ATP binding, motor control
GPR125	G protein-coupled receptor 125	AK014455	6380025	149253878	9.96926186	10.54246965	-0.573207791	0.672120685	1.487828039	0.023656103	g-protein coupled receptor activity, signal transduction, inflammatory response
SELK	selenoprotein K	NM_019979	6130440	111119000	12.70276519	12.26400536	0.438759836	1.35543867	-1.35543867	0.023750956	selenium binding
MMP17	matrix metallopeptidase 17	NM_011846	1850278	118130008	11.40978068	10.6932734	0.716507283	1.643199085	-1.643199085	0.023827129	metalloendopeptidase activity, proteinaceous extracellular matrix
PHACTR1	phosphatase and actin regulator 1	NM_001005740	3190176	54144634	10.89850582	10.33006403	0.568441794	1.482921053	-1.482921053	0.023857499	protein phosphotase inhibitor activity, cell junction and synapse, actin binding
HIST1H2AI	histone cluster 1, H2ai	NM_178182	102480368	30061392	10.65776511	11.76975033	-1.111985224	0.462656953	2.161428686	0.023899124	nucleosome assembly, DNA binding
LDB1	LIM domain binding 1	NM_001113408	5270601	6754519	11.71221711	12.2518761	-0.539658985	0.687933499	1.453628877	0.023945823	Wnt receptor signalling pathway, organismal development, protein binding
2310004I03RIK (Dpp8)	dipeptidylpeptidase 8	NM_028906	104480156		11.924174	11.42593422	0.498239781	1.412489146	-1.412489146	0.023956056	Proteolysis, serine-type endopeptidase activity
RSL1D1	ribosomal L1 domain containing 1	NM_025546	4850309	142385918	10.01810458	10.4004051	-0.382300516	0.767213219	1.303418626	0.024210614	structural constituent of a ribosome, translation
AI850995 (Gfod1)	glucose-fructose oxidoreductase domain containing 1	NM_001033399	106400601	51829580	13.20654998	12.35827336	0.848276616	1.800349017	-1.800349017	0.024416369	metabolic process, electron carrier
FIBCD1	fibrinogen C domain containing 1	NM_178887	1500500	142353012	12.54040893	11.54327645	0.997132481	1.996028722	-1.996028722	0.024419912	signal transduction, receptor binding, integral to membrane
STAC2	SH3 and cysteine rich domain 2	NM_146028	7100022	146149186	10.52777612	9.857914863	0.669861262	1.590919968	-1.590919968	0.024533102	intracellular signalling cascade
SPNB3	spectrin beta 3	NM_021287	6660088	55926126	13.43128635	12.60157028	0.829716074	1.777335544	-1.777335544	0.024552714	actin binding
EG632248	predicted gene, EG632248	XM_001473902	102320725	149266875	9.429775345	10.17977686	-0.750001517	0.594602932	1.681794599	0.024630537	N/A
CX3CL1	chemokine (C-X3-C motif) ligand 1	NM_009142	3990707	114431260	12.99915902	11.82267914	1.176479875	2.260246117	-2.260246117	0.024666151	signal transduction, immune response, cell adhesion, protein binding, cytokine activity, chemokine activity
RAPGEFL1	Rap guanine nucleotide exchange factor (GEF)-like 1	NM_001080925	100380619	124257962	13.04466549	12.03165544	1.013010056	2.018117334	-2.018117334	0.024679244	guanyl-nucleotide exchange factor activity
FGF13	fibroblast growth factor 13	NM_010200	630575	31542808	11.50014783	11.06865541	0.431492424	1.34862797	-1.34862797	0.024726715	MAPKKK cascade, protein binding, growth factor activity
IDH3A	isocitrate dehydrogenase 3 (NAD+) alpha	NM_029573	460332	31560055	10.51283929	10.08591113	0.426928152	1.344368041	-1.344368041	0.024772539	tricarboxylic acid cycle, response to oxidative stress, manganese ion binding, metal ion binding, isocitrate dehydrogenase (NADP+) activity
WIPF3	WAS/WASL interacting protein family, member 3	AK078889	105900577	149255188	13.46482293	12.16808888	1.296734049	2.456721039	-2.456721039	0.024879055	N/A
EIF4A1	eukaryotic translation initiation factor 4A1	NM_144958	1990341	142366509	10.1664051	10.59481943	-0.42841433	0.743078056	1.345753642	0.024997698	translation, RNA binding, ATP binding
COL4A1	collagen, type IV, alpha 1	NM_009931	1740575	33859527	9.326515376	10.19585564	-0.869340264	0.547397115	1.826827312	0.025296502	phosphate transport, cell adhesion, proteinaceous extracellular matrix, extracellular space, structural molecule activity

Table 4.3: Complete list of dysregulated genes in the L100P adult group. Continued next page….

TargetID	Full Gene Name	Accession Number	ProbeID	GI numbers	C57Sal MEAN expression	L100PSal MEAN expression	Difference	anti-log	fold change	Pval	GO Terms
SHANK3	SH3/ankyrin domain gene 3	NM_021423	4070348	10946785	9.868645981	9.460909937	0.407736044	1.326602403	-1.326602403	0.025465175	MAPKKK cascade, intracellular signalling, post synaptic membrane
LEFTY1	left right determination factor 1	NM_010094	3290452	126012509	9.755912081	9.265759779	0.490152302	1.404593147	-1.404593147	0.0255529	negative regulation of cell proliferation, anterior/posterior axis specification
SAP30	sin3 associated polypeptide	NM_021788	5080735	12408289	10.27662024	10.78689456	-0.510274323	0.702088925	1.424320999	0.025612013	negative regulation of transcription, protein binding, sin3 complex
3100002J23RIK	RIKEN cDNA 3100002J23 gene	AK013920	104050139	94390085	9.854060116	9.454578942	0.399481174	1.319033471	-1.319033471	0.025661545	N/A
SYP	synaptophysin	NM_009305	2690168	6678194	15.19090926	13.88818697	1.302723298	2.466941147	-2.466941147	0.025737744	synaptic transmission, endocytosis
PRDX4	peroxiredoxin 4	NM_016764	3800039	146134468	9.428045924	10.01556366	-0.587517734	0.665486943	1.502659083	0.025841026	oxidoreductase activity, peroxidase activity
MOBP	myelin-associated oligodendrocytic basic protein	NM_001039364	7000059	31543261	11.05884095	10.15165759	0.907183361	1.875380531	-1.875380531	0.02584369	structural constituent of myelin sheath, mitochondrion
HIST1H2AD	histone cluster 1, H2ad	NM_178188	101450286	67972409	10.89199134	12.12814483	-1.236153492	0.424502957	2.355696195	0.025907534	nucleosome assembly, DNA binding
EML4	echinoderm microtubule associated protein like 4	NM_001114361	5700373	83921565	9.425861406	9.814976742	-0.389115336	0.763597701	1.309590114	0.025933305	microtubule component
RAB3A	RAB3A, member RAS oncogene family	NM_009001	3940288	31560641	10.59506463	9.876744223	0.718320408	1.645265494	-1.645265494	0.025974156	axonogenesis, exocytosis, small GTPase mediated signal transduction, synaptic vessicle exocytosis and maturation
RQCD1	rcd1 (required for cell differentiation) homolog 1 (S. pombe)	NM_021383	1850139	148540052	9.351013081	9.731886007	-0.380872926	0.767972776	1.302129491	0.025976497	cytokine and chemokine mediated signaling pathway
SEC61B	Sec61 beta subunit	NM_024171	840113	13324683	9.804111303	10.24224794	-0.43813664	0.736087294	1.354853292	0.025996839	integral to membrane, protein transport
MID1IP1	Mid1 interacting protein 1 (gastrulation specific G12-like (zebrafish))	NM_026524	2570739	141801493	11.08315278	10.69519747	0.387955313	1.308537539	-1.308537539	0.026007826	negative regulation of microtubule depolymerisation, nuclear, protein C-terminus binding
IGF2R	insulin-like growth factor 2 receptor	NM_010515	1570402	133778977	9.220759097	9.716007053	-0.495247957	0.709439736	1.409562996	0.026215891	glycoprotein binding, insulin-like growth factor binding, receptor activity, membrane fraction, transport
VSTM2	V-set and transmembrane domain containing 2A	NM_145967.1	5050025	22122404	10.05235803	9.487803403	0.564554626	1.478930879	-1.478930879	0.026234215	extracellular protein, no further info
LITAF	LPS-induced TN factor	NM_019980	6940671	9910577	9.476459177	9.890966649	-0.414506472	0.750276111	1.33284265	0.026259808	protein binding, apoptosis, regulation of transcription
LOC100045304	similar to neurofilament protein	XM_001474024.1	630239	149261713	9.528017304	9.077838365	0.450178939	1.366209699	-1.366209699	0.026350609	N/A
C230067O06RIK (Ptb1)	polypyrimidine tract binding protein 1	AK082600	1049200211		8.358633886	8.977530887	-0.618897001	0.651168582	1.535700627	0.026402532	N/A
EPB4.1L1	erythrocyte protein band 4.1-like 1	NM_001003815	6040288	76056873	10.50537971	10.1052684	0.400111315	1.319609725	-1.319609725	0.026424363	cortical actin cytoskeleton organisation and biogenesis, extrinsic to membrane, actin binding
DGKZ	diacylglycerol kinase zeta	NM_138306	3460041	30794243	9.936746618	9.558211804	0.378534815	1.300020898	-1.300020898	0.026616875	intracellular signalling cascade, activation of protein kinase C activity, metal ion binding, diacyglycerol binding
C630013B14RIK	RIKEN cDNA C630013B14 gene	AK046023	1015002731		10.14719542	9.542224524	0.604970899	1.520948073	-1.520948073	0.026643508	N/A
D630004N19RIK	RIKEN cDNA D630004N19 gene	NM_207274	6760021	46402296	9.440357356	9.047280986	0.39307637	1.313190639	-1.313190639	0.026887975	N/A
KCTD13	potassium channel tetramerisation domain containing 13	NM_172747	2850064	141802609	12.69083788	12.06383128	0.627006599	1.54435733	-1.54435733	0.026984125	potassium ion transport, identical protein binding, voltage gated potassium channel complex
EIF5A	eukaryotic translation initiation factor 5A	NM_181582	1500059	142380083	9.898359139	10.32374134	-0.425382197	0.744641436	1.342928223	0.027194312	translational initiation, apoptosis
ANXA5	annexin A5	NM_009673	1050373	6753059	10.58101589	10.99544613	-0.414430239	0.750315757	1.332772223	0.027201571	Calcium ion binding
HIST2H2AA2	histone cluster 2, H2aa2	NM_178212	7100402	30061398	11.50928756	12.32583974	-0.816552173	0.567797274	1.761191972	0.027417551	nucleosome assembly, DNA binding
ARMC10	armadillo repeat containing 10	NM_026034	5290575	146135072	10.14153112	10.66689243	-0.525361304	0.694786088	1.439293987	0.027901153	regulation of cell growth
LOC674611	similar to zinc finger protein 365	XM_980150.1	5220088	94388323	11.1730491	10.335712	0.837337104	1.786749152	-1.786749152	0.027928705	N/A
CCT3	chaperonin subunit 3 (gamma)	NM_009836	7100403	6753319	9.28014364	9.706891429	-0.426747789	0.743936924	1.344199981	0.027929886	protein folding, protein binding, nucleotide binding
TMEM56	transmembrane protein 56	NM_178936	4540731	118130392	10.28954016	9.778204793	0.511335369	1.425368917	-1.425368917	0.027950903	integral membrane protein
PPP1R9B	protein phosphatase 1, regulatory subunit 9B	NM_172261	3130619	50053702	13.1751468	12.39288133	0.782265461	1.719829398	-1.719829398	0.028004959	actin cytoskeleton organisation and biogenesis, cytoskeletal regulatory protein binding, NS development, cell differentiation
1700019E19RIK	RIKEN cDNA 1700019E19 gene	NM_029601	6940139	21313191	9.506840815	9.900159902	-0.393319087	0.761375955	1.313411586	0.028027374	N/A
ARF3	ADP-ribosylation factor 3	NM_007478	3710446	118131061	12.5289648	11.84757342	0.681391376	1.60368565	-1.60368565	0.028045322	small GTPase mediated signal transduction, protein transport
HIST1H2AK	histone cluster 1, H2ak	NM_178183	102570446	30061370	9.182843528	10.60213616	-1.419292632	0.373895692	2.674543431	0.028222282	nucleosome assembly, DNA binding
LMBRD1	LMBR1 domain containing 1	NM_026719	3170091	123701961	10.50988369	10.01278935	0.497094337	1.411368129	-1.411368129	0.028226199	integral membrane protein
COL4A2	collagen, type IV, alpha 2	NM_009932	2350619	36031079	10.23477261	11.16141977	-0.926647161	0.526079537	1.900853255	0.028239779	phosphate transport, cell adhesion, proteinaceous extracellular matrix, extracellular space, structural molecule activity
PPP2R2C	protein phosphatase 2 (formerly 2A), regulatory subunit B (PR 52), gamma isoform	NM_172994	60279	42475972	10.28260792	9.565506765	0.71710116	1.643875637	-1.643875637	0.02825934	signal transduction, protein phosphate type 2A regulator activity
SCHIP1	schwannomin interacting protein 1	NM_001113419	4560152	148277019	10.78842872	10.30765068	0.480778042	1.395496051	-1.395496051	0.028384896	protein binding (neurofibromin 2 proposed)
SYN1	synapsin I	NM_001110780	4920301	62177179	13.72091856	12.71779457	1.003123988	2.004335459	-2.004335459	0.028394057	neurotransmitter secretion
LRFN2	leucine rich repeat and fibronectin type III domain containing 2	NM_027452	6450411	142388558	10.18758401	9.769638942	0.417945064	1.336023203	-1.336023203	0.028427969	integral membrane protein, protein binding
TSPAN7	tetraspanin 7	NM_019634	870133	84662760	14.1735492	13.23828976	0.935259439	1.91223448	-1.91223448	0.028518243	integral membrane protein
LOC100047385	similar to adenylate cyclase 5	XM_001478039.1	1410402	149267841	11.20550659	10.78643549	0.419071099	1.337066387	-1.337066387	0.028589305	N/A
ATP6V0E2	ATPase, H+ transporting, lysosomal V0 subunit E2	NM_133764	4200458	50659089	12.07798596	11.45352996	0.624456003	1.541629414	-1.541629414	0.028663743	ion transport, proton transport
GALNTL1	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase-like 1	NM_001081421	2850497	124487252	11.86016443	11.23931167	0.620852758	1.537783876	-1.537783876	0.028845306	integral to membrane, transferase activity, calcium ion binding, manganese ion binding
SFRP1	secreted frizzled-related protein 1	NM_013834	6980315	6530147 1	8.472099094	9.290857275	-0.818758181	0.566929725	1.763887049	0.028972723	protein binding , cell differentiation
LOC100044812	similar to cornichon homolog 2 (Drosophila)	XM_001473093.1	5080609	149270282	10.06172136	9.609174399	0.452546957	1.368454016	-1.368454016	0.028993889	N/A
SMARCE1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily e, member 1	NM_020618	4390398	41055974	9.193051768	9.773642076	-0.580590308	0.668690113	1.495461022	0.029066107	chromatin modification, transcription regulation
SULF1	sulfatase 1	NM_172294	430575	26986616	8.689462994	9.174136247	-0.484673253	0.71465891	1.399268918	0.02908469	apoptosis, sulfur metabolism
GAA	glucosidase, alpha, acid	NM_008064	520035	31982354	11.63759512	11.17505901	0.462536113	1.377962009	-1.377962009	0.029147892	neuromuscular process controlling balance/posture, lysosome organisation and biogenesis, locomotory behaviour, glycogen metabolism
UTP11L	UTP11-like, U3 small nucleolar ribonucleoprotein, (yeast)	NM_026031	3460674	146149184	10.0033759	10.41142804	-0.40805214	0.753640217	1.326893095	0.029174683	RNA processing
SULT4A1	sulfotransferase family 4A, member 1	NM_013873	5270161	34328448	14.3200198	13.32037792	0.999641878	1.999503599	-1.999503599	0.029233733	steroid and lipid metabolism
STK16	serine/threonine kinase 16	NM_011494	940239	142387156	10.54041013	9.910101553	0.630308578	1.547896038	-1.547896038	0.029390784	protein amino acid phosphorylation
DKK3	dickkopf homolog 3 (Xenopus laevis)	NM_015814	6450072	31560475	12.75635935	12.13263174	0.623727616	1.540851273	-1.540851273	0.029425896	Wnt receptor signalling pathway, multicellular organismal development
TXNIP	thioredoxin interacting protein	NM_001009935	102640017	13994377	10.19314913	10.60755513	-0.414406	0.750328363	1.332749832	0.029470706	cell cycle, transcription
EFS	embryonal Fyn-associated substrate	NM_010112	110288	141801691	9.364988154	9.903515926	-0.518527772	0.698083844	1.432492685	0.029506509	SH3 domain binding, cell adhesion
ABR	active BCR-related gene	NM_198018	610079	38683819	10.85372004	10.43141603	0.422304009	1.340065956	-1.340065956	0.029568815	GTPase activator activity, brain development
4930488B01RIK	RIKEN cDNA 4930488B01 gene	AK015642	1031200561		11.89668811	11.23426307	0.662425035	1.582740828	-1.582740828	0.029625073	N/A
PPAPDC2	phosphatidic acid phosphatase type 2 domain containing 2	NM_028922	1001101361	142361366	10.72735246	10.30798829	0.419364176	1.337338034	-1.337338034	0.02966839	integral membrane protein, hydrolase activity
HSD11B1	hydroxysteroid 11-beta dehydrogenase 1	NM_001044751	450066	113680660	10.95981176	10.12140427	0.838407493	1.7880753	-1.7880753	0.029675681	steroid metabolic process, lipid metabolism, integral to membrane

Table 4.3: Complete list of dysregulated genes in the L100P adult group. Continued next page….

TargetID	Full Gene Name	Accession Number	ProbeID	GI numbers	C57Sal MEAN expression	L100PSal MEAN expression	Difference	anti-log	fold change	Pval	GO Terms
FBLN2	fibulin 2	NM_001081437	730601	124517703	8.262908121	8.818654083	-0.555745962	0.680305208	1.46992848	0.029762192	calcium ion binding
ACAT2	acetyl-Coenzyme A acetyltransferase 2	NM_009338	2680400	148747460	10.05769085	10.49487255	-0.437181701	0.738576006	1.353956793	0.029775847	acetyl-CoA C-acetyltransferase activity
AGRN	agrin	NM_021604	6770152	42490750	11.20931814	11.6054066	-0.396088461	0.759915831	1.315935211	0.029883517	receptor clustering, synaptic transmission, regulation of synaptic growth at neuromuscular junction
EG667190	predicted gene, EG667190	XR_002058	103850341	94407543	9.429537827	10.20147216	-0.771934335	0.585631745	1.707557707	0.030047115	N/A
ZBTB12	zinc finger and BTB domain containing 12	NM_198886	1850079	40254419	8.876000471	9.647602168	-0.771601697	0.585766788	1.707164046	0.030221584	metal ion binding
RP23-480B19.10	similar to histone 2a	NC_000079.5	2480093	149263930	10.49012479	11.63369157	-1.143566778	0.452639133	2.209265456	0.030289291	N/A
SFRS5	splicing factor, arginine/serine-rich 5 (SRp40, HRS)	NM_001079694	3450176	119226242	9.731170597	10.13467286	-0.40350226	0.756020752	1.322715015	0.03032263	mRNA splice site selection, nucleic acid binding, mRNA processing, transcription
C130065N10RIK	RIKEN cDNA C130065N10 gene	AK048489	102340390		10.9851452	10.35000695	0.635138249	1.553086571	-1.553086571	0.030416187	N/A
SLC25A22	solute carrier family 25 (mitochondrial carrier, glutamate), member 22	NM_026646	2360050	142384143	12.09522768	11.31184035	0.78338733	1.721167293	-1.721167293	0.030437344	mitochondrial transport
MAG	myelin-associated glycoprotein	NM_010758	2370037	118130141	11.31728272	10.38855032	0.928732395	1.903602686	-1.903602686	0.030539078	integral to membrane, cell adhesion, myelin sheath, protein binding, sugar binding
TUBA1C	tubulin, alpha 1C	NM_009448	520041	118130393	9.048838982	9.908530635	-0.859691653	0.551070326	1.814650424	0.030657892	protein polymerisation, microtubule-based movement
LOC631204	similar to SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily e, member 1	XR_030844.1	4920465	149254699	8.312063759	8.802927122	-0.490863363	0.711599123	1.4052856	0.030894222	N/A
ZCCHC5	zinc finger, CCHC domain containing 5	NM_199468	2650021	41054991	8.243160726	8.777725471	-0.534564745	0.690366926	1.448505081	0.031076438	N/A
RPA3	replication protein A3	NM_026632	6700136	146141223	9.138762693	9.543134204	-0.404371611	0.75566532	1.323612308	0.031095772	DNA replication
HYAL2	hyaluronoglucosaminidase 2	NM_010489	1500152	45331201	9.102343819	9.638619594	-0.536275775	0.689548639	1.450224021	0.03133776	metabolic process, carbohydrate metabolic process, receptor activity, anchored to membrane
ROR2	receptor tyrosine kinase-like orphan receptor 2	NM_013846	2680070	47271532	7.954241501	8.48132127	-0.527079769	0.693957988	1.441009423	0.031626595	Wnt receptor signaling pathway, calcium modulating pathway, protein amino acid phosphorylation, transmembrane receptor protein tyrosine kinase activity, transferase activity
4833424O15RIK	RIKEN cDNA 4833424O15 gene	NM_029425	102100538	58037432	9.245157214	8.861141713	0.384015501	1.304968969	-1.304968969	0.031696639	catalytic activity, hydrolase activity
C130092E12	RIKEN clone C130092E12	AK048640	102680397		9.63623781	10.06092827	-0.424690458	0.74499856	1.342284474	0.031788531	N/A
COL1A1	collagen, type I, alpha 1	NM_007742	730020	118131144	8.339505365	9.141139224	-0.801633859	0.573699093	1.743074049	0.031796216	phosphate transport, cell adhesion, extracellular matrix structural component conferring tensile strength
PRKCB	protein kinase C, beta 1	NC_000073.5	870019	6679344	11.24767678	10.46572157	0.78195521	1.71945959	-1.71945959	0.032307467	protein amino acid phosphorylation, cellular calcium ion homeostasis, protein kinase activity, calcium channel regulator activity, metal ion binding, protein binding
PCDH20	protocadherin 20	NM_178685	4200576	142351715	10.95069242	10.21283868	0.737853742	1.667693011	-1.667693011	0.032326795	calcium ion binding, protein binding, cell adhesion
SCN2B	sodium channel, voltage-gated, type II, beta	NM_001014761	10666497	130504904	11.53552626	10.72417553	0.811350726	1.754853659	-1.754853659	0.032331431	ion transport, voltage-gated ion channel activity
4933428A15RIK (Mfap3L)	microfibrillar-associated protein 3-like	NM_027756	101940341	29788791	10.69715179	10.16695456	0.530197231	1.444126608	-1.444126608	0.032333072	protein targeting, integral membrane
CBX7	chromobox homolog 7	NM_144811	3940035	115270984	10.63121782	10.01422568	0.616992138	1.533674302	-1.533674302	0.032357102	chromatin binding, chromatin assembly/disassembly, regulation of transcription
6030405A18RIK	RIKEN cDNA 6030405A18 gene	NM_177854	5290538	31342840	10.42530927	9.776463088	0.648846183	1.567913731	-1.567913731	0.032399448	N/A
SORT1	sortilin 1	NM_019972	1850242	34610210	11.93632097	11.48291342	0.453407551	1.369270568	-1.369270568	0.032413703	vesicle organisation and biogenesis, myotube differentiation, cell differentiation, endocytosis
GGA2	golgi associated, gamma adaptin ear containing, ARF binding protein 2	NM_028758	520451	131890064	9.456350861	9.836666331	-0.38031547	0.768269577	1.301626448	0.032432701	vesicle-mediated transport, protein complex assembly, clathrin adaptor complex
GPD1	glycerol-3-phosphate dehydrogenase 1 (soluble)	NM_010271	2480095	31542905	13.5666598	12.29025738	1.276402421	2.422341754	-2.422341754	0.03267381	glycerol-3-phosphate metabolic process, gluconeogenesis, oxidoreductase activity, NAD binding
SDF4	stromal cell derived factor 4	NM_011341	3440519	118130195	11.19063074	10.75922792	0.431402819	1.34854421	-1.34854421	0.032866722	calcium ion bindging, golgi apparatus
MTDNA_ND4L			104400530		11.50023177	10.32564909	1.212501006	2.317520155	-2.317520155	0.032076043	N/A
1700019B16RIK (Paqr6)	progestin and adipoQ receptor family member VIII	NM_028829	106860538	40254345	11.17320752	10.61083939	0.56236813	1.476691163	-1.476691163	0.032946944	lipid binding, steroid binding
SOX12	SRY-box containing gene 12	NM_011438	1780037	54020661	9.464022258	10.10943753	-0.645416268	0.639308744	1.564189461	0.033045863	regulation of transcription, DNA dependant
PDIA4	protein disulfide isomerase associated 4	NM_009787	70100	86198315	9.467187691	9.888002878	-0.420815187	0.747002416	1.338683757	0.033070971	cell redox homeostasis, isomerase activity
SLC26A4	solute carrier family 26, member 4	NM_011867	5130044	146134337	10.15782148	9.533031149	0.624790326	1.541986705	-1.541986705	0.033096968	transport, integral to membrane, sensory perception of sound
PLA2G7	phospholipase A2, group VII (platelet-activating factor acetylhydrolase, plasma)	NM_013737	4730092	146134462	9.747849444	9.107034178	0.640815266	1.559210019	-1.559210019	0.033112745	lipid catabolic process, inflammatory response, hydrolase activity, catalytic activity
SNRPF	small nuclear ribonucleoprotein polypeptide F	AK008240	3170280	94388566	10.75749814	11.30278741	-0.545289273	0.685253989	1.45931292	0.0331217	nuclear mRNA splicing via spliceosome
ACVR2B	activin receptor IIB	NM_007397	1660546	118131001	9.693068101	10.46647102	-0.773402924	0.585035906	1.709296797	0.033361926	embryonic development, transmembrane receptor protein serine/threonine kinase signaling pathway
SLC6A11	solute carrier family 6 (neurotransmitter transporter, GABA), member 11	NM_172890	6520465	146198773	9.536728251	9.147951372	0.388776879	1.309282919	-1.309282919	0.033370012	neurotransmitter transport
MGLL	monoglyceride lipase	NM_011844	2030446	31982751	11.51616527	10.76673294	0.74943233	1.681131211	-1.681131211	0.033459349	acylglycerol lipase activity, carboxylesterase activity
PRRX1	paired related homeobox 1	NM_001025570	4120193	70909355	7.852484029	8.271939586	-0.419455556	0.74770674	1.337422743	0.033599058	DNA binding, regulation of transcription, multicellular organismal development
CST3	cystatin C	NM_009976	7050452	141803473	15.23920113	14.48473599	0.754465145	1.687006047	-1.687006047	0.03386909	cysteine protease inhibitor activity, endopeptidase inhibitor activity
MT3	metallothionein 3	NM_013603	1450537	7305286	13.43983193	12.19240834	1.247423591	2.374170576	-2.374170576	0.033903925	metal ion binding, negative regulation of neurogenesis, cellular metal ion homeostasis, synaptic vesicle, electron carrier activity
TMEM2	transmembrane protein 2	NM_001033759	1980037	76253921	8.465583271	8.944913555	-0.479330284	0.717310531	1.39409636	0.034045265	integral membrane protein
RIN1	Ras and Rab interactor 1	NM_145495	510593	118130150	10.36290345	9.774422148	0.588481307	1.503663041	-1.503663041	0.034076219	neuropeptide signalling pathway, GTPase activator activity, protein binding
5033413D16RIK	RIKEN cDNA 5033413D16 gene	AK017164	103940142		7.881911416	8.316994014	-0.435082597	0.739651407	1.351988235	0.034081326	N/A
TSPAN2	tetraspanin 2	NM_027533	3940161	70608156	11.02761049	10.3854644	0.64217405	1.560679232	-1.560679232	0.03409338	integral membrane protein
SMC2	structural maintenance of chromosomes 2	NM_008017	4810133	62990165	8.664614768	9.118222912	-0.453608144	0.730214315	1.369460965	0.034101183	chromosome organisation and biogenesis, mitotic chromosome condensation
SLC1A3	solute carrier family 1 (glial high affinity glutamate transporter), member 3	NM_148938	1500070	40254195	13.54002208	12.83641579	0.703606292	1.628570632	-1.628570632	0.034156635	dicarboxylic acid transport, cell projection, transmembrane transport
HIST1H2AB	histone cluster 1, H2ab	NM_175660	2680465	87299610	9.627850499	10.65345015	-1.025599653	0.491206088	2.035805386	0.034221311	nucleosome assembly, DNA binding
PARD6G	par-6 partitioning defective 6 homolog gamma (C. elegans)	NM_053117	4780484	31543459	8.861645709	9.448403304	-0.586757595	0.665837673	1.501867558	0.034428672	cell division, cell cycle, protein binding
EG667728 (Hist1h2al)	histone cluster 1, H2al	AK077055	102450161	149264137	12.3791212	13.00020471	-0.621083511	0.650182436	1.538029858	0.034479863	N/A
YEATS4	YEATS domain containing 4	NM_026570	3130242	146134920	9.347922255	9.728060856	-0.380138601	0.76836377	1.301466882	0.034509397	chromatin modification, regulation of cell growth, transcription
RASL10B	RAS-like, family 10, member B	NM_001013386	100430600	145386572	10.07969419	9.617279923	0.462414272	1.37784564	-1.37784564	0.034538149	small GTPase mediated signal transduction, nucleotide binding
THOC4	THO complex 4	NM_011568	3800731	6755762	10.03164613	10.61800049	-0.586354362	0.6660238	1.501447846	0.03457196	nucleic acid binding, RNA splicing, transport
ENO2	enolase 2, gamma neuronal	NM_013509	2320068	31982744	14.62364904	13.30894124	1.314707797	2.487519434	-2.487519434	0.034619686	glycolysis, phosphopyruvate hydratase complex
ITM2C	integral membrane protein 2C	NM_022417	780647	142386544	14.66742044	14.03522511	0.632195324	1.549921689	-1.549921689	0.034636455	integral membrane protein, ATP binding
LOC665386	similar to prothymosin, alpha (gene sequence 28)	XR_032236.1	101850113	149263919	9.654975902	10.22381411	-0.568836206	0.674159466	1.483328574	0.03466236	N/A
SLC2A3	solute carrier family 2 (facilitated glucose transporter), member 3	NM_011401	1990377	142377808	12.38649122	11.87581999	0.512671233	1.426689349	-1.426689349	0.03468591	glucose transport
YWHAB	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, beta polypeptide	NM_018753	1740176	146134365	13.18017693	12.64878814	0.531388787	1.44531984	-1.44531984	0.034831638	enzyme activator, PKC binding, intracellular signalling cascade, signal transduction

Table 4.3: Complete list of dysregulated genes in the L100P adult group. Continued next page….

TargetID	Full Gene Name	Accession Number	ProbeID	GI numbers	C57Sal MEAN expression	L100PSal MEAN expression	Difference	anti-log	fold change	Pval	GO Terms
ATP1A2	ATPase, Na+/K+ transporting, alpha 2 polypeptide	NM_178405	110278	85861248	12.41025873	11.54206521	0.868193512	1.825375802	-1.825375802	0.034838033	ion transport
FN3K	fructosamine 3 kinase	NM_001038699	1740519	84662726	9.559394238	9.14521096	0.414183278	1.332544099	-1.332544099	0.03487479	fructosamine metabolic process, kinase activity, transferase activity
HIST1H2AN	histone cluster 1, H2an	NM_178184	5050239	30061364	9.706218795	10.96131667	-1.255097875	0.418965143	2.386833407	0.034903259	nucleosome assembly, DNA binding
KCNC4	potassium voltage gated channel, Shaw-related subfamily, member 4	NM_145922	110053	118130234	11.08321182	10.36843306	0.714778758	1.64123151	-1.64123151	0.034945842	cation transport, voltage-gated potassium channel complex
EDNRA	endothelin receptor type A	NM_010332	6900133	93102407	8.674527825	9.122343453	-0.447815628	0.733152067	1.363973512	0.035099577	signal transduction, response to hypoxia, neural crest cell development, negative regulation of cAMP biosynthetic process
9130404D14RIK	RIKEN cDNA 9130404D14 gene	NM_146119	580193	22122640	9.907173913	10.36283644	-0.455662526	0.729175237	1.371412451	0.035143747	N/A
SFRS2	splicing factor, arginine/serine-rich 2 (SC-35)	NM_011358	50707	6755477	10.71866166	11.23205402	-0.513392358	0.70057317	1.427402652	0.035151423	mRNA splice site selection, nucleic acid binding, mRNA processing
SPC24	SPC24, NDC80 kinetochore complex component, homolog (S. cerevisiae)	NM_026282	1940563	146135000	8.207526531	8.803677883	-0.596151352	0.661516317	1.511678509	0.035163206	cell cycle, cell division
6430547I21RIK	RIKEN cDNA 6430547I21 gene	AK032443	100520019		9.252710393	9.694377692	-0.441667299	0.736283206	1.358173039	0.035194921	N/A
FUS	fusion, derived from t(12;16) malignant liposarcoma (human)	NM_139149	4760619	112734868	9.612229213	10.08492923	-0.472700013	0.7206147	1.387704136	0.035276496	postivie regulation of transcription from RNA polymerase II promoter, zinc ion binding, nucleic acid binding, cytoplasmic
CRTAP	cartilage associated protein	NM_019922	3990725	9910169	8.409914817	8.932020703	-0.522105886	0.69635463	1.436049905	0.035322935	spermatogenesis, proteinaceous extracellular matrix
SOX4	SRY-box containing gene 4	NM_009238	2260091	56118238	9.147263986	9.955219615	-0.807955629	0.571190691	1.750728814	0.035331113	regulation of transcription, DNA dependant
PLK1	polo-like kinase 1 (Drosophila)	NM_011121	1780369	126485537	7.992675551	8.706982849	-0.714307298	0.609497709	1.640695258	0.035377774	protein amino acid phosphorylation, transferase activity, protein kinase activity, ATP binding
AWU496U4	expressed sequence AW049624	NM_134096	2320092	54792125	10.87419338	10.40436806	0.469857316	1.384972486	-1.384972486	0.035443673	extracellular region
MT2	metallothionein 2	NM_008630	6860286	123701854	9.899783513	9.457106792	0.442676721	1.359123655	-1.359123655	0.035531953	nitric oxide mediated signal transduction, detoxification of copper ion, cellular zinc ion homeostasis
PSD	pleckstrin and Sec7 domain containing	NM_028627	1170435	51317391	10.36132738	9.720017438	0.641309939	1.559744735	-1.559744735	0.035611806	integral membrane protein
BRD3	bromodomain containing 3	NM_001113573	6130672	142364270	11.11048423	11.64093971	-0.530455484	0.692336117	1.444385141	0.035842811	nucleur
ALDOA	aldolase 1, A isoform	NM_007438	6290672	59709447	15.04606348	14.23575843	0.81030505	1.753582189	-1.753582189	0.036225098	glycolysis, metabolic process
WTIP	WT1-interacting protein	NM_207212	3830195	118130987	9.123809485	9.608482484	-0.484672999	0.714659036	1.399268671	0.03631208	metal ion binding
SIRPA	signal-regulatory protein alpha	NM_007547	130465	110626108	12.92827173	12.22186387	0.706407868	1.631736234	-1.631736234	0.036379237	actin filament organisation, cell motility, phagocytosis
RGS14	regulator of G-protein signaling 14	NM_016758	380086	31980625	9.423523884	8.945043306	0.478480578	1.393275519	-1.393275519	0.036536333	G-protein coupled receptor protein signaling pathway, mitosis, signal transducer activity
ITM2A	integral membrane protein 2A	NM_008409	460576	51556453	8.800386266	9.472558173	-0.672171907	0.627561213	1.593470054	0.036559809	integral membrane protein
FGF1	fibroblast growth factor 1	NM_010197	4780435	122937366	9.247999009	8.861879207	0.386119802	1.306873772	-1.306873772	0.036603587	cell differentiation and proliferation, signal transduction, heparin binding
TMCC2	transmembrane and coiled-coil domains 2	NM_178874	6900762	31341188	9.893187286	9.478244417	0.414942869	1.333245879	-1.333245879	0.03665723	integral membrane protein
NPCD	neuronal pentraxin with chromo domain	NC_000081.5	450438	113374170	10.76482357	10.07977618	0.68504739	1.607754792	-1.607754792	0.036672956	integral to membrane, calcium ion binding, protein binding
TMEM39A	transmembrane protein 39a	NM_026407	450112	27754068	9.035783108	9.462730217	-0.42694711	0.74383415	1.344385707	0.03671209	integral membrane protein
VAMP2	vesicle-associated membrane protein 2	NM_009497	2100100	118130363	13.03786351	12.37939749	0.658466014	1.578403449	-1.578403449	0.036735278	calcium ion dependant exocytosis, synaptic vessicle exocytosis, vesicle mediated transport
CDKN1C	cyclin-dependent kinase inhibitor 1C (P57)	NM_009876	6520577	141803290	9.776171849	11.12454423	-1.348372378	0.392734875	2.546247004	0.036843663	cyclin dependant protein kinase inhibitor activity, negative regulation of transcription from RNA polymerase II promoter, neuron maturation, cell cycle arrest
2310016C16RIK	RIKEN cDNA 2310016C16 gene	NM_027127	3450497	142368412	8.85495376	9.472127759	-0.617173998	0.651946734	1.533867643	0.036987326	response to oxidative stress
APOE	apolipoprotein E	NM_009696	4200671	31981894	13.70425653	12.64083214	1.06342439	2.089886207	-2.089886207	0.037039825	lipid metabolism, lipid transport, regulation of gene expression
LRIG3	leucine-rich repeats and immunoglobulin-like domains 3	NM_177152	3140632	42475967	8.890778221	9.639718674	-0.748940453	0.595040408	1.680558138	0.037147311	integral membrane protein, protein binding
CCK	cholecystokinin	NM_031161	6370368	118130260	12.70174995	11.50486694	1.196883012	2.29243847	-2.29243847	0.037206539	neuron migration, axonogenesis, neuropeptide hormone activity
CISD1	CDGSH iron sulfur domain 1	NM_134007	4610088	146149273	11.45852586	10.93893691	0.519588959	1.433546755	-1.433546755	0.037353437	2iron-2-sulfur cluster binding, mitochondrial membrane, metal ion binding
B3GNT5	UDP-GlcNAc-betaGal beta-1,3-N-acetylglucosaminyltransferase 5	NM_054052	4780528	31542176	7.823434758	8.268209864	-0.444775106	0.734698836	1.361101925	0.037421285	protein amino acid glycosylation, transferase activity
SIRT3	sirtuin 3 (silent mating type information regulation 2, homolog) 3 (S. cerevisiae)	NM_001127351	4010440	11967962	10.82097103	10.38632847	0.434642563	1.35157593	-1.35157593	0.03753356	NAD binding, hydrolase activity, chromatin silencing, protein amino acid deacetylation
H2AFX	H2A histone family, member X	NM_010436	3520082	31981698	9.589378511	10.04224094	-0.45286243	0.730591853	1.368753287	0.037809592	response to DNA damage stimulus, nucleosome assembly, DNA repair, double strand break repair via homologous recombination
TNKS2	tankyrase, TRF1-interacting ankyrin-related ADP-ribose polymerase 2	AK164534	840402	149270766	10.5201753	10.98291722	-0.46274192	0.725605894	1.378158595	0.037862353	NAD+ ADP-ribosyltransferase activity, regulation of multicellular organism growth
CAMK2B	calcium/calmodulin-dependent protein kinase II, beta	NM_007595	2760041	142349559	14.88526257	13.62995113	1.25531144	2.387186761	-2.387186761	0.037869287	regulation of neurotransmitter secretion, protein amino acid phosphorylation, calcium ion transport, calcium and calmodulin dependant protein kinase complex, transferase activity, synapse
SERPINA3N	serine (or cysteine) peptidase inhibitor, clade A, member 3N	NM_009252	5340400	130503300	11.00960941	10.22861282	0.780996592	1.718317451	-1.718317451	0.037937157	endopeptidase inhibitor activity, acute-phase response
5031436O03RIK (Dusp3)	dual specificity phosphatase 3 (vaccinia virus phosphatase VH1-related)	NM_028207	105390008		12.68549893	12.01583119	0.669667736	1.590706573	-1.590706573	0.037941001	protein amino acid dephosphorylation
GPR137	G protein-coupled receptor 137	NM_207220	2630347	46402190	12.69989106	12.25742153	0.442469527	1.358928478	-1.358928478	0.037996463	g-protein coupled receptor activity, signal transduction, inflammatory response
FEN1	flap structure specific endonuclease 1	NM_007999	1770541	47132513	9.46077618	9.894174525	-0.433398346	0.740515405	1.350410799	0.038154649	DNA replication and repair
WBP5	WW domain binding protein 5	NM_011712	7050563	113374169	11.80705706	12.32678426	-0.519727202	0.697503711	1.433684128	0.038221662	mediate protein-protein interactions
MMP14	matrix metallopeptidase 14 (membrane-inserted)	NM_008608	670079	31982190	8.906490278	9.645334706	-0.738844428	0.599219123	1.668838595	0.038292496	proteolysis, cell migration, integral to membrane, metal ion binding
HPCAL4	hippocalcin-like 4	NM_174998	2600484	146198535	13.35506185	12.06920841	1.285853447	2.43826249	-2.43826249	0.038457961	calcium ion binding
NAP1L5	nucleosome assembly protein 1-like 5	NM_021432	360576	148277107	10.97517657	10.16062544	0.814551123	1.758750853	-1.758750853	0.038497241	nucleosome assembly
YAF2	YY1 associated factor 2	NM_024189	940053	142379604	11.26589549	10.74007443	0.525821055	1.439752727	-1.439752727	0.038523496	regulation of transcription
2310051E17RIK	RIKEN cDNA 2310051E17 gene	AK009928	103440731	149270711	10.23983381	9.310219646	0.92961416	1.904766509	-1.904766509	0.038549542	N/A
AHNAK	AHNAK nucleoprotein (desmoyokin)	NM_001039959	6980072	90403606	8.785076751	9.171651537	-0.386574786	0.764943562	1.307285987	0.038684398	intracellular signalling cascade
TMEM10	transmembrane protein 10	NM_153520.1	3840528	23943843	10.23566326	9.5835333	0.652129963	1.571486592	-1.571486592	0.038686813	integral membrane protein
RRM1	ribonucleotide reductase M1	NM_009103	4150433	31982025	8.616344026	9.104108086	-0.48776406	0.713129475	1.402269903	0.038951846	ribonucleoside-diphosphate reductase activity,deoxyribonucleotide biosynthetic process, protein oligomerization
GLUD1	glutamate dehydrogenase 1	NM_008133	4210372	34328124	12.59553077	12.18905228	0.406478497	1.325446553	-1.325446553	0.039131399	transmembrane receptor protein tyrosine kinase signalling pathway, amino acid metabolic process, mitochondrial, glutamate dehydrogenase [NAD(P)+] activity
PTTG1	pituitary tumor-transforming 1	NM_013917	520463	7305426	8.712979946	8.217295353	0.495684593	1.40998967	-1.40998967	0.039187385	response to DNA damage stimulus, homologous chromosome segregation, cysteine protease inhibitor activity, chromosome organisation and biogenesis
KCNAB2	potassium voltage-gated channel, shaker-related subfamily, beta member 2	NM_010598	4850315	31543029	14.29691553	13.18339805	1.11351747	2.16372548	-2.16372548	0.039349948	potassium ion transport, integral to plasma membrane
LOC100046883	similar to CKLF-like MARVEL transmembrane domain containing 3	XM_001476984.1	1940180	149259254	8.98838204	9.616437101	-0.628055061	0.647048132	1.545480082	0.039414687	N/A
WFDC2	WAP four-disulfide core domain 2	NM_026323	2900465	141802067	9.162680932	9.658901868	-0.496220936	0.708961438	1.410513952	0.039577072	endopeptidase and protease inhibitor activity
CHST1	carbohydrate (keratan sulfate Gal-6) sulfotransferase 1	NM_023850	50601	12963844	13.74767203	12.91400877	0.833663252	1.782204948	-1.782204948	0.03959834	inflammatory response, carbohydrate metabolic process, transferase activity, keratan sulfotransferase activity
TMEM132C	transmembrane protein 132C	AK028289	6900132	149254387	8.09575135	8.558973608	-0.463222258	0.725364347	1.378617523	0.039605721	integral membrane protein
LOC433749	similar to Rho family GTPase RhoA	XM_485435.3	580142	94372220	10.70131751	11.1463209	-0.445003386	0.734582592	1.361317312	0.039679282	N/A

Table 4.3: Complete list of dysregulated genes in the L100P adult group. Cont next page……

TargetID	Full Gene Name	Accession Number	ProbeID	GI numbers	C57Sal MEAN expression	L100PSal MEAN expression	Difference	anti-log	fold change	Pval	GO Terms
TRIM37	tripartite motif-containing 37	NM_197987	5270091	37574063	9.336940278	8.676853831	0.660086447	1.580177306	-1.580177306	0.039874832	metal ion binding, protein binding
PPM1E	protein phosphatase 1E (PP2C domain containing)	NM_177167	102480465	118130345	11.9420354	11.36408602	0.57794938	1.492726007	-1.492726007	0.040437896	protein amino acid dephosphorylation, metal ion binding, hydrolase activity, catalytic activity
FAAH	fatty acid amide hydrolase	NM_010173	2850722	142348835	10.09411809	9.593741587	0.500376507	1.414582685	-1.414582685	0.040466343	carbon-nitrogen ligase activity, hydrolase activity
SERPINH1	serine (or cysteine) peptidase inhibitor, clade H, member 1	NM_001111043	6130014	6753303	9.68025867	10.88852804	-1.208269369	0.432787468	2.310602948	0.040548631	response to heat/stress, unfolded protein binding, serine type endopeptidase inhibitor activity
SVOP	SV2 related protein	NM_026805	4230286	72534842	12.9730846	12.46629748	0.506787121	1.420882367	-1.420882367	0.040688384	ion transport, synapse, transmembrane transported activity
GDF1	growth differentiation factor 1	NM_008107	2350132	118130966	11.42276334	10.91454762	0.50821572	1.422290064	-1.422290064	0.040742721	cytokine activity, growth factor activity, extracellular
RASGRP1	RAS guanyl releasing protein 1	NM_011246	1050148	40254575	12.90726932	11.76309626	1.144173064	2.210194085	-2.210194085	0.040774303	small GTPase mediated signal transduction, intracellular signalling cascade, metal ion binding
SNCB	synuclein, beta	NM_033610	3360133	20589946	12.90963649	11.80450106	1.105035424	2.151041596	-2.151041596	0.040877355	dopamine metabolism
LOC100046120	similar to clusterin	XM_001475611.1	5420075	149266011	14.66845638	13.55130912	1.117147251	2.169176207	-2.169176207	0.04118482	N/A
SLC4A1	solute carrier family 4 (anion exchanger), member 1	NM_011403	6290577	6755559	8.08259037	8.728527948	-0.645937578	0.639077332	1.564755858	0.041269518	anion transport, transmembrane
ERAF	erythroid associated factor	NM_133245	2350242	18875417	7.820777686	8.515566937	-0.694789251	0.61779956	1.618647964	0.041429379	protein stabilisation, hemopoiesis, protein folding
PPP1R3B	protein phosphatase 1, regulatory (inhibitor) subunit 3B	NM_177741	6130253	51036623	7.933687132	8.327466585	-0.393779452	0.761133037	1.313830764	0.041523194	carbohydrate/glycogen metabolic process
DLGAP3	discs, large (Drosophila) homolog-associated protein 3	NM_198618	4540113	142351843	9.623590434	9.202126489	0.421463945	1.339285878	-1.339285878	0.041540165	cell-cell signalling, synapse, post synaptic membrane
VASP	vasodilator-stimulated phosphoprotein	NM_009499	7060500	33469092	9.826938386	10.35682902	-0.628890638	0.6930848	1.442819311	0.041587267	actin cytoskeleton organisation and biogenesis, axon guidance, neural tube closure
RPH3A	rabphilin 3A	NM_011286	2190156	31543602	10.00277205	9.48164726	0.521124793	1.435073663	-1.435073663	0.041648874	intracellular protein transport, Rab GTPase binding, metal ion binding
WASF1	WASP family 1	NM_031877	2060484	31982605	14.11703173	13.57644403	0.540487698	1.454464111	-1.454464111	0.041660049	actin filiment polymerisation, cell motility, protein complex assembly
9830169E20Rik (Ptb1)	polypyrimidine tract binding protein 1	AK036745	107000671		8.80626758	9.413874225	-0.607606645	0.656284541	1.523729324	0.041820054	N/A
HBA-X	hemoglobin X, alpha-like embryonic chain in Hba complex	NM_010405	4810168	142350815	7.806601448	9.377588999	-1.570987551	0.336577922	2.971080204	0.041919118	oxygen transport, metal ion binding
PGD	phosphogluconate dehydrogenase	NM_001081274	6840048	124486894	10.61987163	11.00686436	-0.386992736	0.764721989	1.307664765	0.041958512	metabolic processes
ENPP5	ectonucleotide pyrophosphatase/phosphodiesterase 5	NM_032003	1780338	14030778	12.56111545	11.97182785	0.5892876	1.504503643	-1.504503643	0.041995753	nucleotide catabolic process, metabolic process, phosphoric monoester hydrolase activity
CENPL	centromere protein L	NM_027429	103140722	142354931	8.176125913	8.554753573	-0.37862766	0.769168902	1.300104564	0.042012799	chromosome pericentric region
CALU	calumenin	NM_007594	2680403	119672916	9.147011546	9.556656258	-0.409644712	0.752808742	1.328358644	0.042085118	calcium ion binding
SCOC	short coiled-coil protein	NM_001039137	610048	114687673	10.66030118	10.3420887	0.318212477	1.432179654	-1.432179654	0.042089458	N/A
SLC25A18	solute carrier family 25 (mitochondrial carrier), member 18	NM_001081048	4540047	124486669	11.99760483	11.08151175	0.916093078	1.886998241	-1.886998241	0.042258677	symporter activity, transmembrane transport
HNRPK	heterogeneous nuclear ribonucleoprotein K	NM_025279.2	1190348	142350515	8.693083185	9.112401252	-0.419318067	0.747778	1.337295293	0.042310367	mRNA processing, RNA splicing
BUB1B	budding uninhibited by benzimidazoles 1 homolog, beta (S. cerevisiae)	NM_009773	1450288	6753211	8.050320268	8.636752449	-0.586432181	0.665987876	1.501528836	0.04236414	protein amino acid phosphorylation, protein kinase activity, transferase activity
HIST1H2AF	histone cluster 1, H2af	NM_175661	430072	30061378	11.70593946	12.59054132	-0.884601865	0.541636976	1.846255047	0.042384733	nucleosome assembly, DNA binding
ITPR1	inositol 1,4,5-triphosphate receptor 1	NM_010585	3450519	146198791	12.29642464	11.58199211	0.714432526	1.640837678	-1.640837678	0.042547622	voluntary musculoskeletal movement, post-embryonic development, ion transport, in post-synaptic density
MMD2	monocyte to macrophage differentiation-associated 2	NM_175217	4920100	153791750	13.9472035	13.19615069	0.7510528	1.683020559	-1.683020559	0.042554637	integral to membrane, cytolysis
LCT	lactase	NM_001081076	102350136	124487206	10.18060535	9.565816457	0.614786801	1.531333001	-1.531333001	0.042568755	hydrolase activity
CRTAC1	cartilage acidic protein 1	NM_145123	840408	142350981	10.54615593	10.02279067	0.523366262	1.437304028	-1.437304028	0.042659771	extracellular region, calcium ion binding
D15WSU169E	DNA segment, Chr 15, Wayne State University 169, expressed	NM_198420	2350022	38259213	12.12302719	11.71687117	0.406156023	1.32515032	-1.32515032	0.042678249	protein binding, signal transduction
TMEM130	transmembrane protein 130	NM_177735	5700750	118130473	13.16849146	12.77893887	0.389552592	1.309987089	-1.309987089	0.042720866	integral membrane protein, golgi apparatus
EMID2	EMI domain containing 2	NM_024474	360400	19263337	8.336303849	9.421436323	-1.085131474	0.471349312	2.121568814	0.042783163	phosphate transport, proteinaceous extracellular matrix
MAPK6	mitogen-activated protein kinase 6	NM_015806	6760520	76573875	9.073548153	9.45887949	-0.385331338	0.765603146	1.306159733	0.042816029	protein amino acid phosphorylation, cell cycle
CENPE	centromere protein E	NM_173762	2850022	115648100	8.368046762	8.913130764	-0.545084003	0.685351496	1.4591053	0.043117681	attachment of spindle microtubules to kinetochore
SPATA2L	spermatogenesis associated 2-like	NM_030176	1090451	27753957	10.5422572	9.910554707	0.631702493	1.549392319	-1.549392319	0.043237147	N/A
SNCG	synuclein, gamma	NM_011430	3120725	142386371	8.918221048	9.449399848	-0.5311788	0.691989091	1.445109486	0.043329798	axon/ cell soma
MIA1	melanoma inhibitory activity 1	NM_019394	5700273	9506478	8.987279181	9.449705047	-0.462425866	0.725764871	1.377856713	0.043583148	cellular matrix organisation and biogenesis
ADAMTS19	a disintegrin-like and metallopeptidase (reprolysin type) with thrombospondin type 1 motif, 19	NM_175506	2680739	118130275	7.801165554	8.180423609	-0.379258055	0.768832882	1.300672777	0.04360404	proteolysis, metal ion binding
PRDX5	peroxiredoxin 5	NM_012021	1660592	118129958	13.27997154	12.88248325	0.397488295	1.317212671	-1.317212671	0.043824384	peroxisome organisation and biogenesis
SGOL2	shugoshin-like 2 (S. pombe)	NM_199007	2030338	51092284	7.920424587	8.463831204	-0.543406618	0.6861488	1.457409821	0.043832978	meiotic sister chromatid cohesion, centromeric
GRB10	growth factor receptor bound protein 10	NM_010345	6980082	118129993	8.841373044	9.64098766	-0.799614616	0.574502623	1.740636092	0.043868465	signal transduction, receptor activity, SH3/SH2 adaptor activity, MAPKKK cascade
ATP2A1	ATPase, Ca++ transporting, cardiac muscle, fast twitch 1	NM_007504	110309	36031131	7.889247789	8.469071783	-0.579823994	0.669045394	1.494666892	0.043971873	calcium ion transport, regulation of muscle contraction, proton transport
BTG1	B-cell translocation gene 1, anti-proliferative	NM_007569	6200133	40363260	8.835364514	9.279634232	-0.444269718	0.734956252	1.360625203	0.044019326	negative regulation of cell proliferation, regulation of apoptosis
MMP11	matrix metallopeptidase 11	NM_008606	1980619	133893135	7.930291988	8.388847291	-0.458555303	0.727714618	1.374165057	0.044036416	collagen catabolic process, proteolysis
NUDT21	nudix (nucleoside diphosphate linked moiety X)-type motif 21	NM_026623	1240167	109948297	9.240618804	9.747395016	-0.506776212	0.70379335	1.420871623	0.044101287	mRNA processing/cleavage
SCLD0001849.1_2273 (Masp1)	mannan-binding lectin serine peptidase 1	AK031598.1	102810253	26327450	9.535646335	10.14309699	-0.607449658	0.656356958	1.523563529	0.044164272	Innate immune responses
CHAF1A	chromatin assembly factor 1, subunit A (p150)	NM_013733	6040647	142376418	8.344372622	8.739400148	-0.395027526	0.760474867	1.314967849	0.044182717	response to DNA damage stimulus, DNA dependant regulation of transcription
TNNT1	troponin T1, skeletal, slow	NM_011618	3360671	6755840	8.478037475	9.102542158	-0.624504683	0.648642436	1.541681433	0.044186866	structural constituent of cytoskeleton
EBF2	early B-cell factor 2	NM_010095	940324	145966793	7.89526522	8.436261842	-0.542996622	0.686343823	1.456995702	0.044266653	positive regulation of transcription
DNM3OS	dynamin 3, opposite strand	AK020051	102230100	84872167	7.84493097	8.330074303	-0.485143333	0.714426087	1.399724923	0.044330099	N/A
SLC1A1	solute carrier family 1 (neuronal/epithelial high affinity glutamate transporter, system Xag), member 1	NM_009199	520372	118130482	10.85229399	10.24939719	0.602896794	1.518763038	-1.518763038	0.044421743	symporter activity, transmembrane transport
LSAMP	limbic system-associated membrane protein	NM_175548	2810056	118130292	9.984348979	9.578470873	0.405878106	1.324895071	-1.324895071	0.044425357	cell adhesion, protein binding, membrane anchored
NGEF	neuronal guanine nucleotide exchange factor	NM_001111314	3190538	9845276	12.08536481	11.26218486	0.823179957	1.769301554	-1.769301554	0.044462382	regulation of Rho protein signal transduction, NS development, cell differentiation
MCM4	minichromosome maintenance deficient 4 homolog (S. cerevisiae)	NM_008565	2760673	31982196	9.022229876	9.637562977	-0.615333101	0.652779157	1.531911657	0.04450695	DNA replication and repair
RUTBC2 (Sgsm1)	small G protein signaling modulator 1	AK053555	1500647	142352699	10.59613779	10.14854823	0.447589562	1.363759799	-1.363759799	0.044655526	regulation of Rab GTPase activity, Golgi apparatus
SLC30A3	solute carrier family 30 (zinc transporter), member 3	NM_011773	1850451	118130059	11.43087141	10.71103919	0.719832222	1.646990487	-1.646990487	0.044659247	synaptic ion transport
PLEKHG2	pleckstrin homology domain containing, family G (with RhoGef domain) member 2	NM_001083912	1980161	144922715	8.242698814	8.796016514	-0.553317701	0.681451222	1.467456462	0.044978305	regulation of Rho signal transduction, intracellular signalling cascade

Table 4.3: Complete list of dysregulated genes in the L100P adult group. Cont next page……

TargetID	Full Gene Name	Accession Number	ProbeID	GI numbers	C57Sal MEAN expression	L100PSal MEAN expression	Difference	anti-log	fold change	Pval	GO Terms
DIRAS2	DIRAS family, GTP-binding RAS-like 2	NM_001024474	101770551	114326529	13.16239093	12.21144633	0.950944592	1.933137949	-1.933137949	0.045070289	small GTPase mediated signal transduction
MTVR2	mammary tumor virus receptor 2	NM_023166	1240402	31324533	10.33771214	10.77276883	-0.435056687	0.739664691	1.351963954	0.045133306	receptor activity
TSPAN6	tetraspanin 6	NM_019656	2470201	125490374	8.969394659	9.376585655	-0.407190891	0.754090254	1.326101213	0.045192647	membrane protein
CD276	CD276 antigen	NM_133983	3840019	142362675	8.708903526	9.2152667	-0.506363174	0.703994873	1.420464891	0.045274898	regulation of immune response
KRT14	keratin 14	NM_016958	2340402	21489934	7.791572808	8.823637134	-1.032064326	0.489009933	2.044948236	0.045286671	cytoskeletal organisation and biogenesis, cell morphogenesis, protein binding
KIF2C	kinesin family member 2C	NM_134471	6940082	141802894	7.83424433	8.479322097	-0.645077767	0.639458321	1.56382358	0.045376693	microtubule based movement, protein polymerisation
CDCA4	cell division cycle associated 4	NM_028023	5890315	34328499	8.943698242	9.522863398	-0.579165156	0.669350998	1.493984475	0.045456309	cell division, nucleus
CENPA	centromere protein A	NM_007681	5080154	118131195	8.949871836	9.784675136	-0.834803299	0.560659476	1.78361384	0.045523559	nucleosome assembly, DNA binding
DDN	dendrin	NM_001013741	100050048	149267082	13.66870783	12.23946707	1.429240764	2.69304953	-2.69304953	0.045525185	dendrite component
2900056L01RIK	RIKEN cDNA 2900056L01 gene	AK013704	104230427		9.472206813	9.080274624	0.391932188	1.312149578	-1.312149578	0.045607483	N/A
HBP1	high mobility group box transcription factor 1	NM_153198	2510102	29568078	9.280740333	9.674385496	-0.393645162	0.761203889	1.313708475	0.045687964	Wnt receptor signalling pathway, Dna dependant transcription of regulation
LRRC24	leucine rich repeat containing 24	NM_198119	6650193	142375956	10.28012316	9.813613256	0.466509901	1.381762728	-1.381762728	0.045832914	integral membrane protein, protein binding
N28178	expressed sequence N28178	NM_172690	4060253	42476176	9.742676104	9.35649341	0.386182694	1.306930745	-1.306930745	0.045855627	metal ion binding, protein binding
SYT5	synaptotagmin V	NM_016908	4150082	31560793	10.75630626	10.18708759	0.569218669	1.483719804	-1.483719804	0.045904222	neuron projection, synaptic vessicle transport
EN1	engrailed 1	NM_010133	5360168	7106304	7.846356523	8.318609255	-0.472252732	0.720838148	1.387273971	0.045930138	midbrain/hindbrain development, proximal/distal pattern formation, DNA dependant transcription regulation
RASL10A	RAS-like, family 10, member A	NM_145216	6660450	88853579	10.64955092	9.950754989	0.698795929	1.623149549	-1.623149549	0.046025185	small GTPase mediated signal transduction, nucleotide binding
IGF2BP3	insulin-like growth factor 2 mRNA binding protein 3	NM_023670	2810161	31542991	7.867020744	8.428824332	-0.561803588	0.677454713	1.47611343	0.046053008	regulation of translation
A730054J21RIK	RIKEN cDNA A730054J21 gene	AK031124	106980647		10.03353249	10.51536399	-0.481831496	0.716068001	1.396515412	0.046108044	N/A
CDC20	cell division cycle 20 homolog (S. cerevisiae)	NM_023223	3440017	12963586	7.873112905	8.426191917	-0.553079012	0.681563976	1.467213696	0.046284925	cell cycle, cell division
FNBP1L	formin binding protein 1-like	NM_001114665	4730561	23346554	9.684092423	10.41995011	-0.735857587	0.600460943	1.665387253	0.046342013	involved in endocytosis
CDCA3	cell division cycle associated 3	NM_013538	3940397	141801801	8.235837336	9.248720677	-1.012883341	0.495554851	2.017940087	0.046387212	ubiquitin cycle, cell division
ENDOD1	endonuclease domain containing 1	NM_028013	6940048	141803205	11.99640335	11.43340212	0.563001232	1.477339325	-1.477339325	0.046395039	endonuclease activity
GPC5	glypican 5	NM_175500	610735	141802066	9.51923025	9.089176908	0.430053342	1.34728339	-1.34728339	0.046454403	proteinaceous extracellular matrix, protein binding
PALLD	palladin, cytoskeletal associated protein	NM_001081390	4810600	94383582	8.703674461	9.238802659	-0.535128198	0.690097352	1.449070914	0.046468044	Actin filament, cytoskeleton
D9WSU20E (Tmem30a)	transmembrane protein 30A	NM_133718	106620112	31981565	10.95417951	10.51606359	0.438115922	1.354833836	-1.354833836	0.046767122	Integral to membrane
SCAMP2	secretory carrier membrane protein 2	NM_022813	430114	31560286	9.848603143	10.25642867	-0.407825525	0.753758607	1.326684685	0.046865156	Protein transport
TPM2	tropomyosin 2, beta	NM_009416	520735	133892876	8.618714025	9.163958467	-0.545244442	0.685275283	1.459267573	0.046981042	muscle contraction
MEX3A	mex3 homolog A (C. elegans)	NM_001029890	100110152	71274133	8.328965197	9.051797017	-0.72283182	0.605906959	1.650418411	0.047033467	RNA binding
ANXA2	annexin A2	NM_007585	3140402	118131198	10.13127499	11.20880517	-1.077530182	0.473839317	2.110420058	0.047042453	Calcium ion binding
ARHGEF4	Rho guanine nucleotide exchange factor (GEF) 4	NM_183019	610593	56699447	13.84721563	12.83489527	1.012320362	2.017152785	-2.017152785	0.047315393	regulation of Rho protein signal transduction
COL9A1	collagen, type IX, alpha 1	NM_007740	4570369	31982451	7.828124127	8.437635417	-0.60951129	0.655418686	1.52574228	0.047453859	cell adhesion, phosphate transport
ZER1	zer-1 homolog (C. elegans)	NM_178694	580524	31542270	13.27245632	12.82021775	0.452238562	1.368161522	-1.368161522	0.0476477	protein binding
CAMTA2	calmodulin binding transcription activator 2	NM_178116	3990075	118130545	14.16818136	13.21149902	0.956682342	1.940841551	-1.940841551	0.047657991	regulation of transcription DNA dependant
PLXNB2	plexin B2	AK003544	106020176	149267067	9.67798273	10.08726214	-0.409279408	0.752999385	1.328022332	0.04777319	positive regulation of axonogenesis
VASH2	vasohibin 2	NM_144879	2030059	21450246	8.891422907	9.915669625	-1.024246718	0.491666949	2.033697138	0.047848792	N/A
PRC1	protein regulator of cytokinesis 1	NM_145150	5890204	21553110	8.094686826	8.890876338	-0.796189512	0.57586817	1.736508547	0.048017893	cytokinesis, cell cycle, cell division
BZRAP1	benzodiazapine receptor associated protein 1	NM_172449	5290487	27369600	12.0463666	11.513669	0.5326976	1.446631629	-1.446631629	0.048042254	receptor activity (benzodiazapine) , mitochondrion
ACTA1	actin, alpha 1, skeletal muscle	NM_009606	840538	133893192	9.34794156	9.976102543	-0.628160983	0.647000628	1.545593554	0.0480473	cytoskeleton organisation and biogenesis, muscle contraction
TEF	thyrotroph embryonic factor	NM_017376	1050050	118130107	11.9109672	11.37905833	0.531908868	1.445840961	-1.445840961	0.04809068	positive regulation of transcription from RNA polymerase II protmoter
MAP2K4	mitogen-activated protein kinase kinase 4	NM_009157	5130133	145966829	11.87270696	11.49208209	0.380624863	1.301905617	-1.301905617	0.048139695	protein amino acid phosphorylation, MAPKKK cascade, transferase activity
PANX3	pannexin 3	NM_172454	6560592	86262154	7.962939649	8.543655728	-0.580716079	0.668631821	1.495591399	0.048187765	gap junction, membrane
NUSAP1	nucleolar and spindle associated protein 1	NM_001042652	940048	111118995	7.839826012	8.295710257	-0.455884245	0.729063184	1.371623232	0.048310977	establishment of mitotic spindle localisation, cytokenisis after mitosis
SIPA1L3	signal-induced proliferation-associated 1 like 3	NM_001081028	4730167	124486629	9.951506191	9.46810037	0.483405821	1.398040176	-1.398040176	0.048384344	protein binding, GTPase activator activity
CHAF1B	chromatin assembly factor 1, subunit B (p60)	NM_028083	3940142	146141116	7.856085989	8.464846014	-0.608760026	0.655760076	1.524947976	0.048400954	response to DNA damage stimulus, DNA dependant regulation of transcription
NDRG1	N-myc downstream regulated gene 1	NM_008681	1340040	118150657	10.60114914	10.16068513	0.440464009	1.357040718	-1.357040718	0.048440503	mast cell activation
NDG2	Nur77 downstream gene 2	NM_175329	6660400	34013513	14.15468695	13.05393808	1.100748868	2.144659877	-2.144659877	0.048443883	mitochondrial
SLC17A7	solute carrier family 17 (sodium-dependent inorganic phosphate cotransporter), member 7	NM_182993	7000037	33859826	13.37154409	12.05290777	1.318636325	2.494302301	-2.494302301	0.048681941	sequestering of neurotransmitters, excitatory synapse
P4HB	prolyl 4-hydroxylase, beta polypeptide	NM_011032	6110056	118130201	11.23917444	11.81666452	-0.577490074	0.670128619	1.492250848	0.048909909	cell redox homeostasis
MSH6	mutS homolog 6 (E. coli)	NM_010830	4480064	6754743	9.34926274	9.775899782	-0.426637043	0.743994034	1.3440968	0.049126313	DNA damage response, signal transduction resulting in induction of apoptosis, determination of adult lifespan
SF3A3	splicing factor 3a, subunit 3	NM_029157	6660603	124249078	10.77727476	11.19285423	-0.415579471	0.749718303	1.333834316	0.049171569	mRNA processing, RNA splicing
MAL	myelin and lymphocyte protein, T-cell differentiation protein	NM_010762	4280487	141801061	9.715240128	9.241412143	0.473827985	1.388789538	-1.388789538	0.049238552	induction of apoptosis, integral to membrane
SST	somatostatin	NM_009215	6590142	6678034	13.49616348	12.65270285	0.843460629	1.794349135	-1.794349135	0.049608479	regulation cell migration, hormone activity
PTGIS	prostaglandin I2 (prostacyclin) synthase	NM_008968	6620411	141802458	8.261509978	8.727876671	-0.466366693	0.723785097	1.381625574	0.049646257	Fatty acid biosynthesis
CDC2A	cell division cycle 2 homolog A (S. pombe)	NM_007659	1570411	141803421	7.796123062	8.275958023	-0.479834961	0.717059648	1.394584122	0.049660529	mitotic cell cycle G2/M transition DNA damage checkpoint, cell cycle, cell division
MKI67	antigen identified by monoclonal antibody Ki 67	NM_001081117	3440750	149258042	7.851071783	8.382249296	-0.531177513	0.691989709	1.445108197	0.049664297	cell proliferation, meiosis
PRNP	prion protein	NM_011170	3140687	13173472	15.81949561	15.05282232	0.766673294	1.701342137	-1.701342137	0.049808406	protein homo-oligomerisation, cellular copper ion homeostasis, response to oxidative stress
TNNT2	troponin T2, cardiac	NM_011619	2450364	6755842	8.547387585	9.069305739	-0.521918154	0.69644525	1.43586305	0.049845503	sarcomere organisation
DPP6	dipeptidylpeptidase 6	NM_010075	5810368	24429567	10.2814881	9.857391137	0.424096961	1.341732398	-1.341732398	0.049993359	proteolysis

Table 4.3: Complete list of dysregulated genes in the L100P adult group. Columns show Target (gene) ID, full gene name, accession number and probe ID, log transformed expression values for both comparable groups, difference between groups, fold-change, p-value and GeneOntology terms for each gene.

Nine genes showed overlap with the previously published schizophrenia CNV study by Walsh et al [83] and six of these findings have the same presumed directionality as the CNV study (ie over/under-expression corresponding to gain/loss of copy number). Two genes were found to overlap with a second CNV study [84] and again followed the same directionality (Table 4.2). Comparison with a microarray study of the F22 family described above (Christoforou *et al*, unpublished) also showed two genes of overlap with the current study following the same directionality.

Gene	Study	Gain/Loss	Function
GP1BB	Xu, Gogs and Karayiorgou (2008) Schizophrenia CNVs	Down-regulated, CNV deletion	hemostasis, cell adhesion, protein binding
ARMC10	Walsh et al (2008) Schizophrenia CNVs (cases)	Up-regulated, CNV duplication	regulation of cell growth
DLG2	Walsh et al (2008) Schizophrenia CNVs (cases) Xu, Gogs and Karayiorgou (2008) Schizophrenia CNVs	Down-regulated, CNV deletion	preotein binding, sensory perception of pain, synaptic transmission
EMID2	Walsh et al (2008) Schizophrenia CNVs (cases)	Up-regulated, CNV duplication	phosphate transport, proteinaceous extracellular matrix
HBP1	Walsh et al (2008) Schizophrenia CNVs (cases)	Up-regulated, CNV duplication	Wnt receptor signalling pathway, DNA dependant regulation of transcription
LARGE	Walsh et al (2008) Schizophrenia CNVs (cases)	Down-regulated, CNV duplication	transferase activity and biosynthetic processes
MLL5	Walsh et al (2008) Schizophrenia CNVs (cases)	Up-regulated, CNV duplication	metal ion binding, regulation of transcription
PPP1R3B	Walsh et al (2008) Schizophrenia CNVs (controls)	Up-regulated, CNV deletion	carbohydrate/glycogen metabolic process
SLC1A3	Walsh et al (2008) Schizophrenia CNVs (cases)	Down-regulated, CNV deletion	Dicarbolic acid transport, cell projection, transmembrane transport
SLC26A4	Walsh et al (2008) Schizophrenia CNVs (cases)	Down regulated, CNV duplication	ion transport, integral to membrane, sensory perception of sound
PLXNB2	Christoforou et al (unpublished) F22 family affected vs MIC	Down-regulated, Down-regulated	negative regulation of transcription
SORT1	Christoforou et al (unpublished) F22 family affected vs MIC	Down-regulated, Down-regulated	implicated in modulation of dopamine signals

Table 4.2: Overlap with previous studies. This table shows the overlaps between genes from the current microarray study and large genome wide studies previously published from individuals with schizophrenia and other major mental illness. All genes with overlap were found within the L100P non-drug treated (saline) group. Column 1 shows the gene name, column 2 the comparative previously published evidence, column 3 the direction of change (with this study first and comparative study second) and column 4 the gene function/location.

GOTree analysis of the differentially expressed gene list did not find any categories to be overenriched in the sample dataset (table 4.2, figure 4.2).

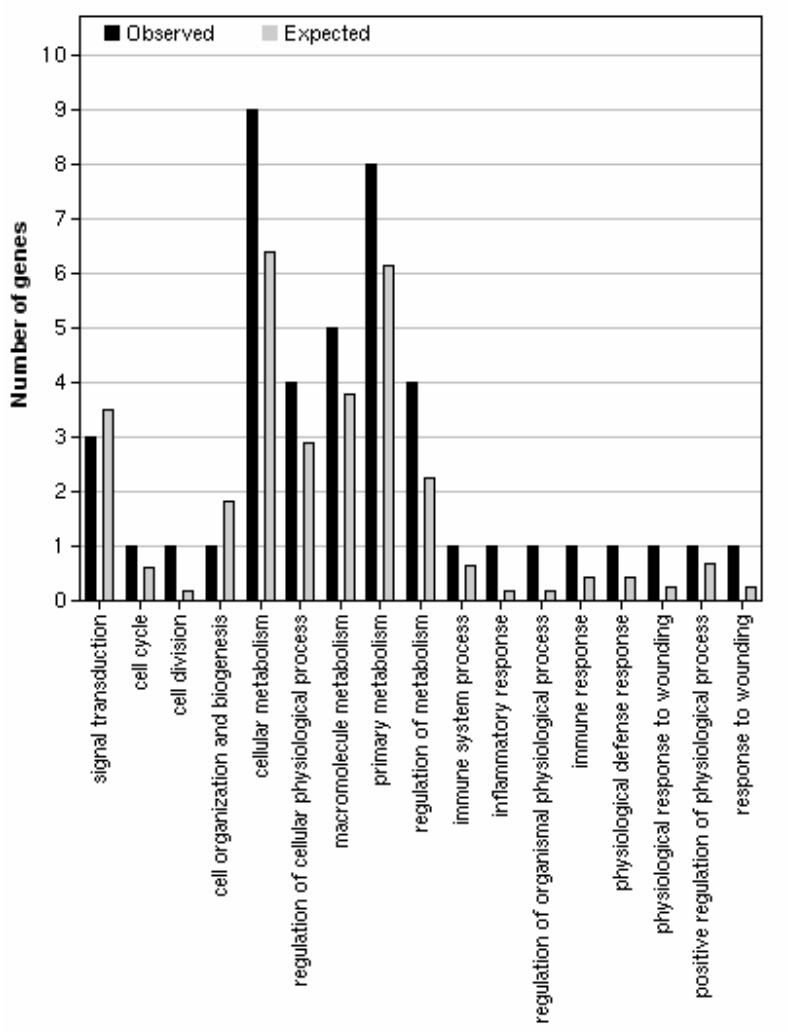

Figure 4.2: Bar graph of GOTree categories for L100Padult vs C57BL6adult. For each biological process GOTM calculates the expected number of genes within a dataset and graphs this against the observed genes within each category. A p-value is then calculated to determine the significance of over/under-enrichment for each category. For this dataset none of the categories were significantly overenriched.

Pathway analysis of the L100P dysregulated genes using IPA gave three statistically significant high scoring networks. The top network for the L100P mutant adults was "neurological disorder" (IPA score 45. Scores greater than three have at least 99.9%

confidence that the genes are not grouped by random chance alone. For full score definition see Chapter 2 section 7.2) It contained 32 of the L100P dysregulated genes and had a high network score, with low probability these genes clustered by chance (fig 4.3). Calcium signalling was cited as the top canonical pathway. A canonical pathway is a core pathway established for a given molecule in the cell, in which molecular interactions occur in a linear and stepwise manner. Network 2 (IPA score 43) “cell cycle” contained 28 L100P dysregulated genes while network 3 (IPA score 41) “developmental disorder”, also contained 28 L100P dysregulated genes, including one gene, *Bex2*, whose expression level was corrected toward the wild-type level by rolipram treatment (Figure 4.3).

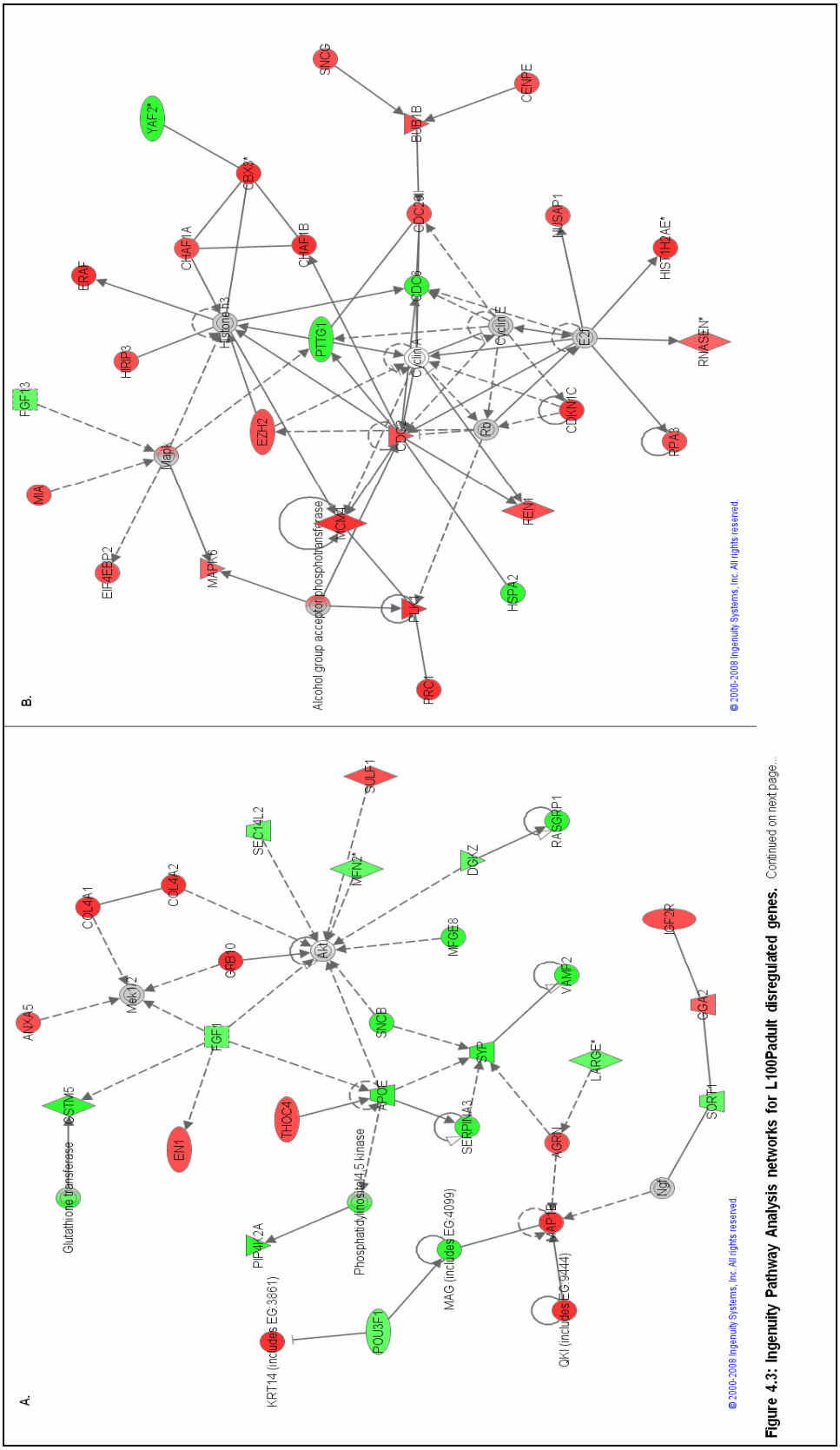

Figure 4.3: Ingenuity Pathway Analysis networks for L100Padult disregulated genes. Continued on next page...

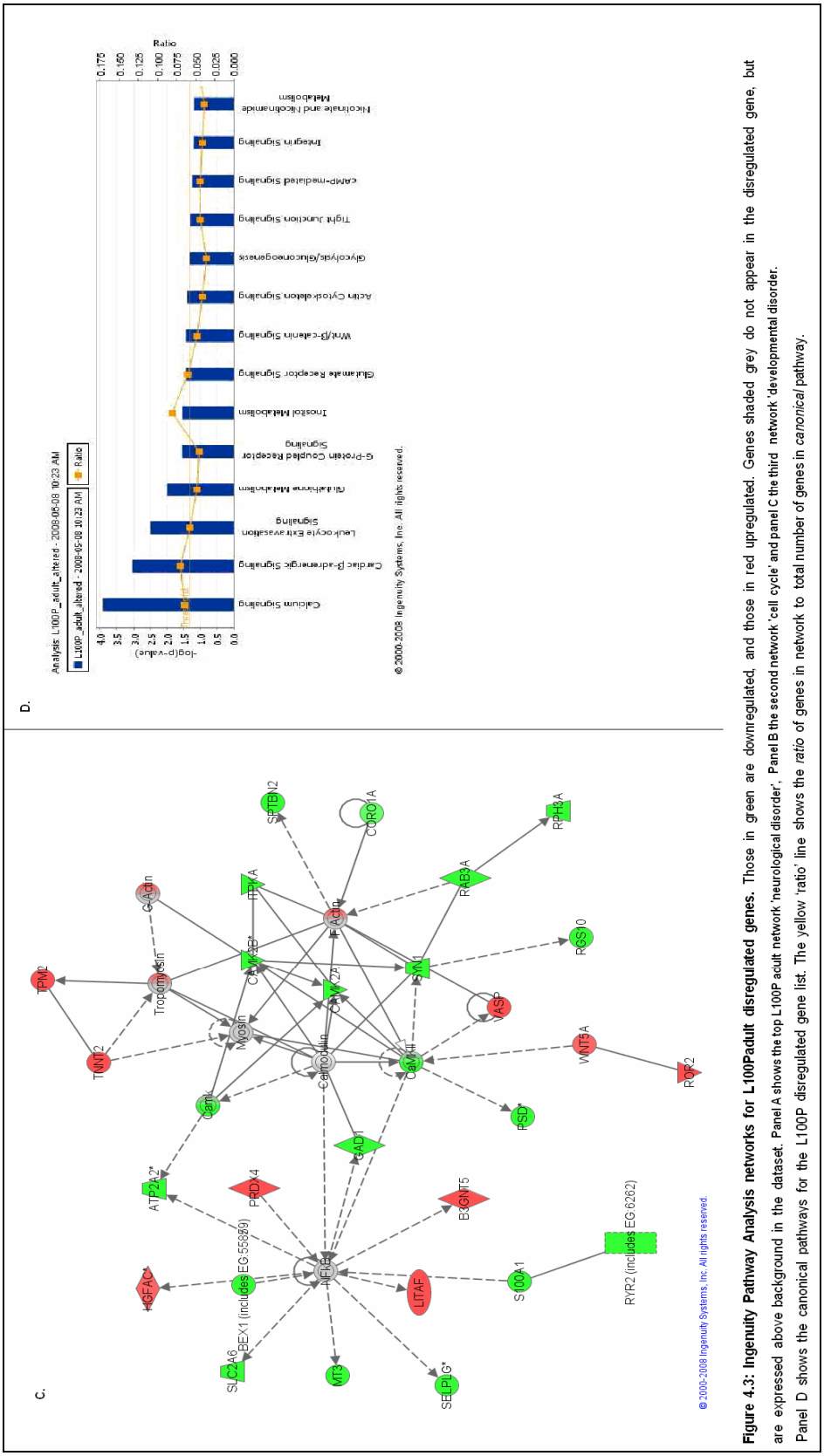

Figure 4.3: Ingenuity Pathway Analysis networks for L100Padult disregulated genes. Those in green are downregulated, and those in red upregulated. Genes shaded grey do not appear in the disregulated gene, but are expressed above background in the dataset. Panel A shows the top L100P adult network 'neurological disorder'. Panel B the second network 'cell cycle' and panel C the third network 'developmental disorder. Panel D shows the canonical pathways for the L100P disregulated gene list. The yellow 'ratio' line shows the *ratio* of genes in network to total number of genes in *canonical* pathway.

4.2.4 Differential Expression between the Q31L adult ENU mutant and the C57BL/6J adult controls

The differentially expressed genes in the Q31L non-drug treated (saline) adult mice are shown in table 4.4. For this comparison, the top network contained only two genes from the array study. This was, however, expected as fewer genes were differentially expressed in this group. The small number of genes was insufficient to create canonical pathways.

TargetID	Full Gene Name	Accession Number	ProbeID	GI numbers	C57Sal MEAN expression	Q31LSal MEAN expression	Difference	anti-log	fold change	Pval	GO Terms
WDFY1	WD repeat and FYVE domain containing 1	AC083910	1580463	141803291	8.16741092	8.757048349	-0.58963743	0.66450989	1.504868507	4.94E-11	metal ion binding
TTC27	tetratricopeptide repeat domain 27	AI118251	1090280	141802682	9.71574372	9.296720764	0.419022951	1.33702177	-1.33702177	0.00020638	binding protein
NDN	necdin	AA103206	5670075	116642896	14.3384393	14.78147057	-0.4430313	0.73558741	1.359457738	0.00231655	transcription, sensory perception of pain, neuron migration, nerve growth factor receptor signalling pathway, axonogenesis, axonal extension, gamma tubulin binding, protein binding
DEFB11	defensin beta 11	AJ437646	2370440	21361235	8.20971136	8.60357547	-0.39386411	0.76108837	1.313907863	0.00380448	defense response to bacterium, extracellular region
D230046H12RIK	RIKEN full-length enriched library, clone:D230046 H12	AK084437	103440746		8.70280728	9.108606158	-0.40579887	0.75481821	1.324822312	0.00409229	N/A
ERDR1	erythroid differentiation regulator 1	AJ007909	5890184	19111169	10.4399921	9.785104997	0.654887149	1.57449279	-1.57449279	0.00787428	biological process unknown, cellular component
GLRX1	glutaredoxin	NM_053108	3520471	31981457	11.2325449	10.8030977	0.429447233	1.34671748	-1.34671748	0.02446397	electron carrier, oxidation reduction, cell redox homeostasis, transferase activity
2900040C04RIK	RIKEN cDNA 2900040C04 gene	XM_001479776	104670008	149265804	9.17218803	9.680789314	-0.50860129	0.70290358	1.422670229	0.03026606	N/A
1500015O10RIK	RIKEN cDNA 1500015O10 gene	AI848847	3130368	142360223	9.67682021	10.22315296	-0.54633275	0.68475854	1.460368798	0.03171745	extracellular localisation

Table 4.4: Dysregulated genes in the Q31L adult group compared to C57BL/6J controls. Table shows the probe target (gene) ID, full name, accession and probe number, log transformed expression data for both Q31L and C57BL/6J, difference in expression, fold-change and p-value, and Gene Ontology terms for each target as available.

4.2.5 Differential Expression between the Embryonic ENU mutant and the C57BL/6J embryonic controls

In total 33 genes were found to be differentially expressed in the L100P embryo, and 15 in the Q31L embryo (table 4.5).

A.

TargetID	Full Gene Name	Accession Number	ProbeID	GI numbers	C57Embryo MEAN	L100PEmbryo MEAN	Difference	Anti-Log	Fold change	Pval	GO Term
XPOT	exportin, tRNA (nuclear export receptor for tRNAs)	NM_001081056	7050184	124486685	10.1304175	10.55105317	-0.4206357	0.747095	1.33851721	0.001879	tRNA binding, transport
B930001L07RIK		AK046898	105360086		9.74405783	10.21983945	-0.4757816	0.719077	1.39067144	0.002347	N/A
LOC100044475	similar to SH2/SH3 adaptor protein	XM_001472256.1	2510010	149233947	8.91208789	8.428745339	0.4833426	1.397979	-1.3979789	0.003778	N/A
1700074P13RIK	RIKEN cDNA 1700074P13 gene	NM_028550	5670161	133893180	8.61638177	9.270723223	-0.6543415	0.635365	1.57389735	0.005846	peptidase activity, extracellular
DDX46	DEAD (Asp-Glu-Ala-Asp) box polypeptide 46	NM_145975	104010026	78126141	9.64773246	10.07069778	-0.4229653	0.74589	1.34068037	0.005947	mRNA processing, RNA splicing
DMRT2	doublesex and mab-3 related transcription factor 2	NM_145831	2630068	146149265	9.73617191	8.252955409	1.4832165	2.795713	-2.7957135	0.006649	regulation of transcription, sex differentiation
PRMT6	protein arginine N-methyltransferase 6	NM_178891	105360154	126432553	9.21329295	8.051993644	1.1612993	2.236588	-2.2365877	0.007013	metabolic, methytransferase activity
OLFR1024	olfactory receptor 1024	NM_001005230	240768	52421325	10.7194293	11.60361441	-0.8841852	0.541793	1.84572185	0.00798	G protein coupled receptor signalling, sensory perception of smell
PDE4D	phosphodiesterase 4D, cAMP specific	NM_011056	2470528	118130222	9.10233961	10.05725226	-0.9549127	0.515873	1.93846226	0.008658	cAMP catabolism and signal transduction
D130004A15RIK	RIKEN cDNA D130004A15 gene	AA673478	102680156		10.6249345	9.796985458	0.827949	1.77516	-1.77516	0.010632	N/A
PRM3	protamine 3	NM_013638	4810577	114326494	10.8513901	10.2432383	0.6081518	1.524305	-1.5243052	0.011449	mitotic chromosome condensation, multicellular organismal development
4833441J24RIK (Ttll3)	tubulin tyrosine ligase-like family, member 3	NM_133923.2	3870142	31981612	11.5577473	12.1534821	-0.5957348	0.661707	1.51124213	0.011762	tubulin-tyrosine ligase activity
ST7	Suppression of tumorigenicity 7	NM_001083315	60600	134032044	8.55960596	7.87687233	0.6827336	1.605178	-1.6051784	0.012927	N/A
2810425O13RIK (Cab39l)	calcium binding protein 39-like	AK013161	100050131		8.50245034	8.048390027	0.4540603	1.36989	-1.3698902	0.020424	N/A
GPR103	G protein-coupled receptor 103	NM_198192	5910066	67514557	8.32067691	7.816408443	0.5042685	1.418404	-1.418404	0.021428	G protein coupled receptor signalling
VTI1-PENDING	vesicle transport through interaction with t-SNAREs homolog 1A	AK035895	102480446		10.0640549	10.63507811	-0.5710232	0.673139	1.48557683	0.021964	Intracellular protein transport
2610028A01RIK	RIKEN cDNA 2610028A01 gene	NM_028228	3360154	141803006	8.6185554	8.998918295	-0.3803629	0.768244	1.30166923	0.02272	telomere maitentance, negative regulation of cell cycle
LOC634350	similar to Heterogeneous nuclear ribonucleoprotein A1 (Helix-destabilizing protein) (Single-strand binding protein) (hnRNP core protein A1) (HDP-1) (Topoisomerase-inhibitor suppressed)	XR_032088.1	106400184	149264511	9.0386739	8.658329425	0.3803445	1.301653	-1.3016526	0.023089	N/A
A530020A01RIK		AK040719	101990112		10.22484	9.736278009	0.488562	1.403046	-1.4030457	0.024389	N/A
A330021D07RIK (Sfrs12)	splicing factor, arginine/serine-rich 12	AK039311	104120142		10.301405	9.910595525	0.3908095	1.311129	-1.3111289	0.025087	RNA splicing
CORO1B	coronin, actin binding protein 1B	NM_011778	6350592	142388278	8.97989403	8.291533925	0.6883601	1.611451	-1.6114508	0.026296	actin binding, cytoskeleton
CDA	cytidine deaminase	NM_026176	1980605	58037288	10.2689374	9.885938721	0.3829987	1.30405	-1.3040495	0.028372	cytadine metabolic process
TIFP39 (Pth2)	parathyroid hormone 2	NM_053256.1	6200341	16716586	10.3936639	10.81882951	-0.4251656	0.744753	1.34272665	0.028857	neuropeptide signalling pathway, cAMP biosythetic process
RHOX5	reproductive homeobox 5	NM_008818	7040066	91206463	9.6904849	10.1417005	-0.4512156	0.731426	1.36719175	0.030743	germ cell programmed cell death, regulation of transcription
KAZALD1	Kazal-type serine peptidase inhibitor domain 1	NM_178929	380603	142353193	9.21401701	8.631246497	0.5827705	1.497723	-1.4977227	0.036018	cell differentiation, regulation of cell growth
LOC100048816	similar to Brain-specific angiogenesis inhibitor 2	XM_001471973.1	630113	149253134	11.071713	11.48021917	-0.4085062	0.753403	1.32731079	0.03781	N/A
8030474H12RIK	RIKEN cDNA 8030474H12 gene	AK033243	106590100		10.0518096	10.44893816	-0.3971286	0.759368	1.3168843	0.038771	N/A
NUMB	numb gene homolog (Drosophila)	NM_010949	2450735	6754911	8.9574875	8.238238048	0.7192495	1.646325	-1.6463253	0.039978	neuroblast proliferation, axonogenesis
GSTM7	glutathione S-transferase, mu 7	NM_026672	5570647	113679873	10.1226274	10.70938039	-0.586753	0.66584	1.50186273	0.044844	metabolic function
LOC100044089	similar to potassium channel regulatory protein KChAP	XM_001472067.1	240739	149251667	11.8650966	11.08283436	0.7822622	1.719826	-1.7198256	0.045559	N/A
LOC380654		XM_354561.1	106040451	38090681	10.3366136	9.605119545	0.731494	1.660358	-1.6603576	0.045768	N/A
CSTF1	cleavage stimulation factor, 3' pre-RNA, subunit 1	NM_024199	4610129	133892406	9.19208747	9.908860543	-0.7167731	0.608457	1.64350184	0.0475	mRNA polyadenylation, mRNA cleavage
CDH11	cadherin 11	NM_009866	1230133	114687887	13.834166	14.46403296	-0.629867	0.646236	1.5474223	0.048901	homophillic cell adhesion

B.

TargetID	Full Gene Name	Accession Number	ProbeID	GI numbers	C57Embryo MEAN	Q31LEmbryo MEAN	Difference	Anti-log	Fold Change	Pval	GO Terms
PDE4D	phosphodiesterase 4D, cAMP specific	NM_011056	2470528	118130222	9.10233961	9.604965579	-0.502626	0.705821	1.41679004	0.006917	cAMP catabolism and signal transduction
COPE	coatomer protein complex, subunit epsilon	NM_021538	2340088	10946971	9.84406121	9.403464202	0.440597	1.357166	-1.3571658	0.011323	retrograde vessicle mediated transport
CORO1B	coronin, actin binding protein 1B	NM_011778	6350592	142388278	8.97989403	8.193499346	0.7863947	1.724759	-1.7247589	0.012092	actin binding, cytoskeleton
B230325K09RIK (Upf2)	UPF2 regulator of nonsense transcripts homolog	AK045943	106510632		8.31303118	7.934164257	0.3788869	1.30032	-1.3003202	0.01273	N/A
6720403M19RIK	RIKEN cDNA 6720403M19 gene	AK032699	106110204		9.33534638	9.748086526	-0.4134601	0.75082	1.33187635	0.020745	N/A
NUMB	numb gene homolog (Drosophila)	NM_010949	2450735	6754911	8.9574875	8.187421071	0.7700664	1.705348	-1.7053483	0.022887	neuroblast proliferation, axonogenesis
5830454D03RIK (Hook3)	hook homolog 3	XM_134108.3	101990068		8.66109273	8.280826169	0.3802666	1.301582	-1.3015823	0.026482	microtubule cytoskeletal organisation and biogenesis
LOC100040996	similar to diacylglycerol kinase, eta	XM_001475622.1	106380735	149264874	10.2147697	9.763039116	0.4517306	1.36768	-1.3676799	0.027185	N/A
C030011I16RIK	RIKEN cDNA C030011I16 gene	AK021065	103850546		8.89719092	8.481621099	0.4155698	1.333825	-1.3338254	0.032505	N/A
PRM3	protamine 3	NM_013638	4810577	114326494	10.8513901	10.46501797	0.3863722	1.307102	-1.3071024	0.033568	mitotic chromosome condensation, multicellular organismal development
EG619517	predicted gene, EG619517	NM_001038500	6180484	84370325	8.82983346	8.350123831	0.4797096	1.394463	-1.394463	0.034275	N/A
2310022M17RIK	RIKEN cDNA 2310022M17 gene	NM_001024919	6660037	67906186	8.87892781	8.288349641	0.5905782	1.50585	-1.5058501	0.035561	N/A
A430073A17RIK (Itgav)	integrin alpha V	AK079805	100110411		8.67302344	8.158694765	0.5143287	1.428329	-1.4283293	0.036841	Integrin mediated cell pathway, apoptotic cell clearance
A830061P03RIK	RIKEN cDNA A830061P03 gene	AK043972	104560114		7.98683637	8.392808215	-0.4059718	0.754728	1.32498116	0.040775	N/A
2410131K14RIK	RIKEN cDNA 2410131K14 gene	NM_001081236	106110139	124486988	11.7518215	11.22238172	0.5294398	1.443369	-1.4433686	0.041549	extracellular localisation

Table 4.5: Dysregulated genes in the Embryonic sample groups. Panel A shows the genes dysregulated in the L100P embryo compared to C57BL/6J controls while Panel B shows the genes dysregulated in the Q31L embryo compared to the C57BL/6J control. Each table shows probe target ID (gene symbol), full gene name, accession number, Illumina probe ID, GI number, log transformed expression data, expression difference, anti-log of expression difference (to correct for negative values), fold-change, p-value and GeneOntology terms where available.

For the L100P E13.5 embryos, there were nine genes in the top network (IPA score 22) and Notch signaling was the top canonical pathway. The Q31L E13.5 embryos had a lower but still significant top network (IPA score 10), with only four genes interconnecting. Three of these genes were also present in the L100P E13.5 embryo top network which added weight to taking them forward for further analysis. Consistent with this, the top canonical pathway for this group was also Notch signaling (figure 4.4).

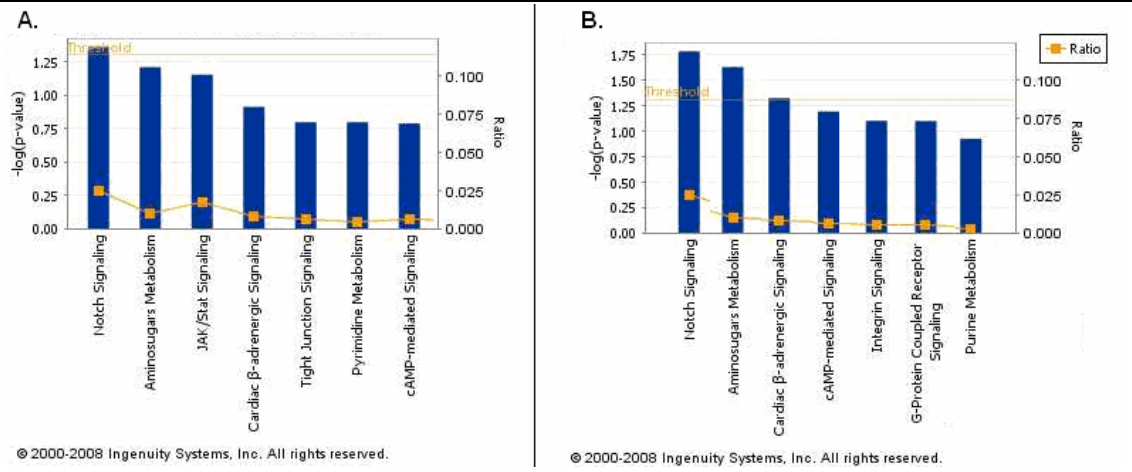

Figure 4.5: Ingenuity Pathway Analysis of Embryonic dysregulated gene lists. Panels A and B show the top seven cananocal pathways for L100P E13 and Q31L E13 respectively. In both cases Notch signaling is the top canonical pathway, and in the case of the L100P, the only one to meet the significance threshold. *Threshold* bar shows cut-off point of significance $P < 0.05$. The yellow 'ratio' line shows the *ratio* of genes in network to total number of genes in *canonical* pathway.

4.2.6 The effect of Drug Treatment on gene expression in adult ENU mutant mice

In total there were 150 genes dysregulated in the drug treatment groups compared to their drug naïve (saline) counterparts (table 4.6). While many of these genes were unique to the drug treatment groups, some had previously been seen to be differentially expressed in the non-drug treated (saline) groups compared to C57BL/6J animals. Treatment of the L100P adult with rolipram showed the most extensive gene expression rescue, with seven genes returning to levels not significantly different to wild-type (fold change +/-1.3,

p<0.05). Clozapine showed no correction of expression and the bupropion treated Q31L animals only had one gene show corrected expression.

TargetID	Full Gene Name	Accession Number	ProbeID	GI numbers	L100PSal MEAN	L100PCloz MEAN	Difference	Anti-log	Fold change	Pval	GO Terms
TMEM68	transmembrane protein 68	NM_028097	1770458	133892404	11.03173	11.526861	-0.495126	0.7094995	1.409444183	0.0309667	integral membrane protein
SCCPDH	saccharopine dehydrogenase (putative)	NM_178653	6180014	142382199	11.49912	11.909879	-0.410758	0.7522278	1.329384532	0.0339684	metabolic processing
JMJD3	jumonji domain containing 3	NM_001017426	4200398	62945393	8.42191	8.1131191	0.3087913	1.2386695	-1.238669478	0.0479497	negative regulation of transcription from RNA polymerase II promoter

Table 4.6.1: Dysregulated genes in Clozapine treated L100P adults. Table shows the genes that are dysregulated in L100P adult mice treated with anti-psychotic Clozapine, compared to L100P drug naïve adult mice. Table shows probe target ID (gene symbol), full gene name, accession number, Illumina probe ID, GI number, log transformed expression data, expression difference, anti-log of expression difference (to correct for negative values), fold-change, p-value and GeneOntology terms where available.

TargetID	Full Gene Name	Accession Number	ProbeID	GI numbers	L100PSal MEAN	L100PRoli MEAN	Difference	anti-log	Fold change	Pval	GO Terms
Sep-15	selenoprotein	NP_444332.1	1780451	110825968	9.875723	9.3441287	0.5315943	1.4455257	-1.445525701	0.0117927	de novo' post-translational protein folding
DUSP1	dual specificity phosphatase 1	NM_013642	6860121	145301574	10.69992	11.203473	-0.50355	0.7053688	1.417698034	0.0142112	hydrolase activity, MAP kinase tyrosine/serine/threonine phosphotase activity, protein amino acid dephosphorylation
LOC381892 (Gm1693)	gene model 1693	XP_358673.5			9.829038	9.2664142	0.5626241	1.4769531	-1.476953148	0.0149717	N/A
NARFL	nuclear prelamin A recognition factor-like	NM_026238			11.08363	11.632196	-0.54857	0.6836977	1.462634799	0.0168551	nucleur protein, no further info
DRG1	developmentally regulated GTP binding protein 1	NM_007879	1660372	6681224	10.10316	9.6208952	0.4822748	1.3969446	-1.396944554	0.0216486	GTP binding
5730589K01RIK (Tox4)	TOX high mobility group box family member 4	NP_075923.1	4810286	12963678	9.729534	9.3384237	0.3911104	1.3114024	-1.311402389	0.0229936	DNA dependant transcription regulation
LOC433749	similar to Rho family GTPase RhoA	XP_485435.1	6801142	94372220	11.14632	10.328014	0.818307	1.7633355	-1.763335464	0.0230017	N/A
PNRC1	proline-rich nuclear receptor coactivator 1	NM_001033225.2			9.911573	9.4013662	0.5102063	1.4242539	-1.424253876	0.0242396	DNA dependant transcription regulation
BCAS3	breast carcinoma amplified sequence 3	NM_138681	4050647	141801917	11.93908	12.35235	-0.413272	0.7509185	1.331702359	0.0244673	nucleur protein, no further info
5730437N04RIK	RIKEN cDNA 5730437N04 gene	NM_027457	2320022	146134493	11.38893	10.798137	0.5907941	1.5060755	-1.506075537	0.0249887	integral membrane protein
EG633120	predicted gene, EG633120	XR_001899	1.02E+08	149270297	14.67498	14.290895	0.3840836	1.3050306	-1.305030588	0.0255865	N/A
D130059O18RIK (Sfrs6)	splicing factor, arginine/serine-rich 6	AK051608	1.06E+08		8.594064	8.1151058	0.4789586	1.3937372	-1.393737236	0.0256353	RNA splicing, mRNA processing
RANBP9	RAN binding protein 9	NM_019930	4670685	9910285	9.741117	9.3139606	0.4271558	1.3445812	-1.344581155	0.0274091	GTPase binding
PPP1R8	protein phosphatase 1, regulatory (inhibitor) subunit 8	NM_146154	6290292	22122684	9.829209	9.4257741	0.4034349	1.3226533	-1.322653298	0.0277948	nucleur mRNA splicing, multicellular organismal development
LOC100048803	hypothetical protein LOC100048803 (related to Psmb6)	XM_001471617.1	2690711	149252027	12.76081	12.322307	0.4385019	1.3551964	-1.355196374	0.0282016	N/A
ARHGAP21	Rho GTPase activating protein 21	NM_001081364	2100168	124486928	12.35369	12.834003	-0.480316	0.7168208	1.395048876	0.0303338	signal transduction
HNRPK	heterogeneous nuclear ribonucleoprotein K	NP_079555.1	1190348	142350515	9.112401	8.4683571	0.6440441	1.5627036	-1.56270356	0.0333425	mRNA processing
LOC100047707	hypothetical protein LOC100047707	XM_001478713.1	6860458	149270676	10.811	11.237779	-0.426783	0.7439188	1.344232775	0.0341739	N/A
PDIA3	protein disulfide isomerase associated 3	NM_007952	2120270	112293263	11.85888	11.056581	0.8022961	1.7438731	-1.743873129	0.0367079	isomerase activity, cell redox homeostasis, positive regulation of apoptosis, extracellular space and endoplasmic reticulum
TSC22D1	TSC22 domain family, member 1	NM_207652	1340739	46559425	9.315611	8.8739925	0.4416186	1.3581272	-1.358127202	0.0369411	DNA dependant transcription regulation
A330021D07RIK (Sfrs12)	splicing factor, arginine/serine-rich 12	AK039311			8.670335	8.2402871	0.4300481	1.3472785	-1.34727845	0.0372455	RNA splicing, mRNA processing
LOC213411	similar to protein tyrosine phosphatase 4a1	XM_124808.2	1.01E+08	38085321	9.661538	9.0818922	0.5796458	1.4944823	-1.494482324	0.0374054	N/A
LOC100046590	hypothetical protein LOC100046590				10.08605	9.7069102	0.3791433	1.3005693	-1.300569349	0.0377819	N/A
RABGEF1	RAB guanine nucleotide exchange factor (GEF) 1	NM_019983	4010497	31982702	11.28478	11.710272	-0.425495	0.744583	1.343033615	0.0382093	endocytosis
TCEA2	transcription elongation factor A (SII), 2	NM_009326	6370546	133892334	10.86471	11.558607	-0.693896	0.6181813	1.617648422	0.0385003	DNA dependant transcription regulation
BEX2	brain expressed X-linked 2	NM_009749	5670433	94408188	13.62231	14.112721	-0.490413	0.7118213	1.404846966	0.0385106	nucleus and cytoplasm, no further info
LOC100046136	hypothetical protein LOC100046136 (related to Bat1a)	XM_001476489.1	4850692	149269277	9.038196	8.4078774	0.6303185	1.5479067	-1.547906684	0.0393093	N/A
SMU1	smu-1 suppressor of mec-8 and unc-52 homolog (C. elegans)	NM_021535	5910128	141803589	8.979609	8.5954841	0.3841253	1.3050683	-1.305068305	0.0395526	N/A
E230027K01RIK	RIKEN cDNA A630054L15 gene	AK054197	1.03E+08		8.689457	8.3055942	0.3838632	1.3048313	-1.304831259	0.0408115	N/A
PCBP1	poly(rC) binding protein 1	NP_035995.1	1050088	47271536	10.98543	10.177571	0.8078559	1.7506078	-1.750607779	0.0409273	mRNA processing, translation activator activity, ribonucleoprotein complex
EG620120	proline rich region 18	NM_178774			13.62317	12.553784	1.0693868	2.0985411	-2.09854115	0.0418426	N/A
PRR18	transmembrane emp24 domain trafficking protein 2	NM_019770	2030064	40254340	9.981408	10.605038	-0.623629	0.6490362	1.540746097	0.0419844	endoplasmic reticulum
TMED2	nuclear import 7 homolog (S. cerevisiae)	NM_025391	2370025	55742882	9.800014	9.1891558	0.6108584	1.5271676	-1.527167611	0.0424324	vesicle mediated transport
NIP7	gamma-aminobutyric acid receptor associated protein	NM_019749	4730039	13928673	9.20324	8.8158594	0.3873806	1.3080164	-1.308016375	0.0427315	ribosome biogenesis and assembly, translation
GABARAP	hairy and enhancer of split 6 (Drosophila)	NM_019479	1450286	40254379	12.50124	12.089163	0.4120762	1.3305993	-1.330599308	0.0427489	microtubule cytoskeletal organisation and biogenesis
HES6	hairy and enhancer of split 6 (Drosophila)	NP_062352.1			9.147603	8.6725686	0.4750342	1.3899512	-1.389951192	0.0430652	negative regulation of neuron differentiation, nervous system development
1190009E20RIK (Usp46)	ubiquitin specific peptidase 46	NP_808229.1	1.02E+08		11.27601	11.704582	-0.428571	0.7429975	1.345899575	0.0434853	ubiquitin dependant protein catabolic process
LOC545208	predicted gene, EG545208		1.06E+08	94402964	9.577063	9.1008907	0.4761725	1.3910483	-1.391048269	0.0435046	N/A
LOC100048153	WW domain binding protein 4	NM_018765	5107722	149254483	13.42924	12.710539	0.7186998	1.6456983	-1.645698267	0.0450747	N/A
WBP4	protein phosphatase 2 (formerly 2A), catalytic subunit, beta isoform	NM_017374	2230019	124286863	10.27399	9.794667	0.4793278	1.394094	-1.394094008	0.0463283	mRNA processing and RNA splicing
PPP2CB	protein phosphatase 2 (formerly 2A), catalytic subunit, beta isoform	NP_059070.1	5570593	119672926	11.72031	11.202101	0.5182098	1.432177	-1.432177013	0.0464249	protein amino acid dephosphorylation
LOC100046623	FMS-like tyrosine kinase 1	NM_010228	1.06E+08	149271925	9.607672	9.0254561	0.582216	1.4971471	-1.497147133	0.0465197	N/A
FLT1	SH3-binding domain glutamic acid-rich protein like	NM_019989	3830167	118129969	9.730684	10.13245	-0.401766	0.7669312	1.321123964	0.0484093	transmembrane receptor protein tyrosine kinase signalling pathway
SH3BGRL	SH3-binding domain glutamic acid-rich protein like	NP_064373.1	2060110	31543698	9.64113	9.249967	0.3911633	1.3114505	-1.311450477	0.0498063	N/A

Table 4.6.2: Dysregulated genes in Rolipram treated L100P adults. Table shows the genes that are dysregulated in L100P adult mice treated with the PDE4 inhibitor and anti-depressant rolipram, compared to L100P drug naïve adult mice. Each table shows probe target ID (gene symbol), full gene name, accession number, Illumina probe ID, GI number, log transformed expression data, expression difference, anti-log of expression difference (to correct for negative values), fold-change, p-value and GeneOntology terms where available.

TargetID	Full Gene Name	Accession Number	ProbeID	GI numbers	L100PSal MEAN	L100PRoll MEAN	Difference	anti-log	Fold change	Pval	GO Terms
Sep-15	selenoprotein	NP_444332.1	1780451	110825968	9.875723	9.3441287	0.5315943	1.4455257	-1.445525701	0.0117927	de novo' post-translational protein folding
DUSP1	dual specificity phosphatase 1	NM_013642	6860121	145301574	10.69992	11.203473	-0.50355	0.7053688	1.417698034	0.0142112	hydrolase activity, MAP kinase tyrosine/serine/threonine phosphotase activity, protein amino acid dephosphorylation
LOC381892 (Gm1693)	gene model 1693	XP_358673.5			9.829038	9.2664142	0.5626241	1.4769531	-1.476953148	0.0149717	N/A
NARFL	nuclear prelamin A recognition factor-like	NM_026238			11.08363	11.632196	-0.54857	0.6836977	1.462634799	0.0168551	nucleur protein, no further info
DRG1	developmentally regulated GTP binding protein 1	NM_007879	1660372	6681224	10.10316	9.6206852	0.4822748	1.3969446	-1.396944554	0.0216486	GTP binding
5730589K01RIK (Tox4)	TOX high mobility group box family member 4	NP_075923.1	4810286	12963678	9.729534	9.3384237	0.3911104	1.3114024	-1.311402389	0.0229936	DNA dependant transcription regulation
LOC433749	similar to Rho family GTPase RhoA	XP_485435.1	580142	94372220	11.14632	10.326014	0.818307	1.7633355	-1.763335464	0.0230017	N/A
PNRC1	proline-rich nuclear receptor coactivator 1	NM_001033225.2			9.911573	9.4013662	0.5102063	1.4242539	-1.424253876	0.0242396	DNA dependant transcription regulation
BCAS3	breast carcinoma amplified sequence 3	NM_138681	4050647	141801917	11.93908	12.35235	-0.413272	0.7509185	1.331702359	0.0244673	nucleur protein, no further info
5730437N04RIK	RIKEN cDNA 5730437N04 gene	NM_027457	2320022	146134493	11.38893	10.798137	0.5907941	1.5060755	-1.506075537	0.0249687	integral membrane protein
EG633120	predicted gene, EG633120	XR_001899	1.02E+08	149270297	14.67498	14.290895	0.3840836	1.3050306	-1.305030588	0.0255865	N/A
D130059O18RIK (Sfrs6)	splicing factor, arginine/serine-rich 6	AK051608	1.06E+08		8.594064	8.1151058	0.4789686	1.3937372	-1.393737236	0.0256353	RNA splicing, mRNA processing
RANBP9	RAN binding protein 9	NM_019930	4670685	9910285	9.741117	9.3139606	0.4271568	1.3445812	-1.344581155	0.0274091	GTPase binding
PPP1R8	protein phosphatase 1, regulatory (inhibitor) subunit 8	NM_146154	6290292	22122684	9.829209	9.4257741	0.4034349	1.3226533	-1.322653298	0.0277948	nucleur mRNA splicing, multicellular organismal development
LOC100048803	hypothetical protein LOC100048803 (related to Psmb6)	XM_001471617.1	2690711	149252027	12.76081	12.322307	0.4385019	1.3551964	-1.355196374	0.0282016	N/A
ARHGAP21	Rho GTPase activating protein 21	NM_001081364	2100168	124486928	12.35369	12.834003	-0.480316	0.7168208	1.395048876	0.0303338	signal transduction
HNRPK	heterogeneous nuclear ribonucleoprotein K	NP_079555.1	1190348	142350515	9.112401	8.4683571	0.6440441	1.5627036	-1.56270356	0.0333425	mRNA processing
LOC100047707	hypothetical protein LOC100047707	XM_001478713.1	6860458	149270676	10.811	11.237779	-0.426783	0.7439188	1.344232775	0.0341739	N/A
PDIA3	protein disulfide isomerase associated 3	NM_007952	2120270	112293263	11.85888	11.056581	0.8022951	1.7438731	-1.743873129	0.0367079	isomerase activity, cell redox homeostasis, positive regulation of apoptosis, extracellular space and endoplasmic reticulum
TSC22D1	TSC22 domain family, member 1	NM_207652	1340739	46559425	9.315611	8.8739925	0.4416186	1.3581272	-1.358127202	0.0369411	DNA dependant transcription regulation
A330021D07RIK (Sfrs12)	splicing factor, arginine/serine-rich 12	AK039311			8.670335	8.2402871	0.4300481	1.3472785	-1.34727845	0.0372455	RNA splicing, mRNA processing
LOC213411	similar to protein tyrosine phosphatase 4a1	XM_124808.2	1.01E+08	38085321	9.661538	9.0818922	0.5796458	1.4944823	-1.494482324	0.0374054	N/A
LOC100046590	hypothetical protein LOC100046590				10.08605	9.7069102	0.3791433	1.3005693	-1.300569349	0.0377819	N/A
RABGEF1	RAB guanine nucleotide exchange factor (GEF) 1	NM_019983	4010497	31982702	11.28478	11.710272	-0.425495	0.744583	1.343033615	0.0382093	endocytosis
TCEA2	transcription elongation factor A (SII), 2	NM_009326	6370546	133892334	10.86471	11.558607	-0.693898	0.6181813	1.617648422	0.0385003	DNA dependant transcription regulation
BEX2	brain expressed X-linked 2	NM_009749	5670433	94408188	13.62231	14.112721	-0.490413	0.7118213	1.404846966	0.0385106	nucleus and cytoplasm, no further info
LOC100046136	hypothetical protein LOC100046136 (related to Bat1a)	XM_001476489.1	4850692	149269277	9.038196	8.4078774	0.6303185	1.5479067	-1.547906684	0.0393093	N/A
SMU1	smu-1 suppressor of mec-8 and unc-52 homolog (C. elegans)	NM_021535	5910128	141803589	8.979609	8.5954841	0.3841253	1.3050683	-1.305068305	0.0395526	N/A
E230027K01RIK	RIKEN cDNA A630054L15 gene	AK054197	1.03E+08		8.689457	8.3055942	0.3838632	1.3048313	-1.304831259	0.0408115	N/A
PCBP1	poly(rC) binding protein 1	NP_035995.1	1050088	47271536	10.98543	10.177571	0.8078559	1.7506078	-1.750607779	0.0409273	mRNA processing, translation activator activity, ribonucleoprotein complex
EG620120	proline rich region 18	NM_178774			13.62317	12.553784	1.0693868	2.0985411	-2.09854115	0.0418426	N/A
PRR18	transmembrane emp24 domain trafficking protein 2	NM_019770	2030064	40254340	9.981408	10.605038	-0.623629	0.6490362	1.540746097	0.0419844	endoplasmic reticulum
TMED2	nuclear import 7 homolog (S. cerevisiae)	NM_025391	2370025	55742882	9.800014	9.1891558	0.6108584	1.5271676	-1.527167611	0.0424324	vesicle mediated transport
NIP7	gamma-aminobutyric acid receptor associated protein	NM_019749	4730039	13928673	9.20324	8.8158594	0.3873806	1.3080164	-1.308016375	0.0427315	ribosome biogenesis and assembly, translation
GABARAP	hairy and enhancer of split 6 (Drosophila)	NM_019479	1450286	40254379	12.50124	12.089163	0.4120762	1.3305993	-1.330599308	0.0427489	microtubule cytoskeletal organisation and biogenesis
HES6	hairy and enhancer of split 6 (Drosophila)	NP_062352.1			9.147603	8.6725686	0.4750342	1.3899512	-1.389951192	0.0430652	negative regulation of neuron differentiation, nervous system development
1190009E20RIK (Usp46)	ubiquitin specific peptidase 46	NP_808229.1	1.02E+08		11.27601	11.704582	-0.428571	0.7429975	1.345899575	0.0434853	ubiquitin dependant protein catabolic process
LOC545208	predicted gene, EG545208		1.06E+08	94402964	9.577063	9.1006907	0.4761725	1.3910483	-1.391048269	0.0435046	N/A
LOC100048153	WW domain binding protein 4	NM_018765	510722	149254483	13.42924	12.710539	0.7186998	1.6456983	-1.645698267	0.0450747	N/A
WBP4	protein phosphatase 2 (formerly 2A), catalytic subunit, beta isoform	NM_017374	2230019	124286863	10.27399	9.794667	0.4793278	1.394094	-1.394094008	0.0463283	mRNA processing and RNA splicing
PPP2CB	protein phosphatase 2 (formerly 2A), catalytic subunit, beta isoform	NP_059070.1	5570593	119672926	11.72031	11.202101	0.5182098	1.432177	-1.432177013	0.0464249	protein amino acid dephosphorylation
LOC100046623	FMS-like tyrosine kinase 1	NM_010228	1.06E+08	149271925	9.607672	9.0254561	0.582216	1.4971471	-1.497147133	0.0465197	N/A
FLT1	SH3-binding domain glutamic acid-rich protein like	NM_019989	3830167	118129969	9.730684	10.13245	-0.401766	0.7569312	1.321123964	0.0484093	transmembrane receptor protein tyrosine kinase signalling pathway
SH3BGRL	SH3-binding domain glutamic acid-rich protein like	NP_064373.1	2060110	31543698	9.64113	9.249967	0.3911633	1.3114505	-1.311450477	0.0498063	N/A

Table 4.6.3: Dysregulated genes in Bupropion treated Q31L adults. Continued on next page…

TargetID	Full Gene Name	Accession Number	ProbeID	GI numbers	Q31LSal MEAN	Q31LBupr o MEAN	Difference	Anti.log	Fold change	Pval	GO Terms
NDUFS2	NADH dehydrogenase (ubiquinone) Fe-S protein 2	NM_153064	4860020	142376353	10.12648	9.6117194	0.5147576	1.4287639	-1.428763918	0.0193363	4 iron, 2 sulfur cluster binding; electron carrier, NADH dehydrogenase activity, mitochondrial respiratory chain, oxidation reduction
TMEM50A	transmembrane protein 50A	NM_027935	2690440	62990167	11.48193	10.670941	0.8109853	1.7544092	-1.754409244	0.0197731	Transmembrane protein with domain of unknown function
LOC100046853	similar to heterogeneous nuclear ribonucleoprotein K	XM_001476909.1	1.06E+08	149264564	11.81413	11.386095	0.4280342	1.3453991	-1.345399142	0.0202722	N/A
ARHGAP21	Rho GTPase activating protein 21	NM_001081364	2100168	124486928	12.47699	12.917517	-0.440525	0.7368666	1.357097817	0.0211711	signal transduction
RPL26	ribosomal protein L26	NM_009080	5220592	130500915	13.19334	12.729004	0.4643382	1.3796843	-1.379684343	0.0216422	ribosome biogenesis and assembly, translation
LOC100043141	similar to ribosomal protein L10a	XM_001479716.1	4850066	149233916	11.38344	10.837876	0.5455643	1.4595912	-1.459591159	0.0220137	N/A
PFDN5	prefoldin 5	NM_027044			13.13072	12.721708	0.409015	1.3277689	-1.327778941	0.0223772	protein folding, unfolded protein binding
AHDC1	AT hook, DNA binding motif, containing 1	NM_146155	5050747	124244059	9.781351	10.164393	-0.383042	0.7668192	1.304088321	0.0226634	DNA binding
RNF130	ring finger protein 130	NM_021540	4230731	133892292	10.08086	9.6518533	0.4290018	1.3463018	-1.346301764	0.0230888	integral membrane protein, metal ion binding, apoptosis
TSPAN3	tetraspanin 3	NM_019793	4560647	31981021	9.94548	9.4751647	0.470315	1.3854119	-1.385411906	0.0237594	integral membrane protein
BCAS1	breast carcinoma amplified sequence 1	NM_029815	2030711	30023813	11.98745	12.419425	-0.431976	0.7412457	1.349080335	0.0239285	protein homodimerization
N4WBP5-PENDING (Ndfip1)	Nedd4 family interacting protein 1	AI-050560			11.57676	10.845388	0.7313689	1.6602136	-1.660213614	0.0240022	N/A
A230065H16RIK	RIKEN cDNA A230065H16 gene	NM_001101503	1.06E+08	94393146	8.887943	8.2913001	0.5966425	1.5121932	-1.512193196	0.0246144	N/A
HERPUD1	homocysteine-inducible, endoplasmic reticulum stress-inducible, ubiquitin-like domain member 1	NM_022331	7100440	11612514	9.511376	9.0707795	0.4405968	1.3571656	-1.357165641	0.0246158	protein modification process, response to stress, response to unfolded protein, endoplasmic reticulum membrane
MOG	myelin oligodendrocyte glycoprotein	NM_010814	610364	113199770	10.23825	9.7277829	0.510466	1.4245093	-1.424509264	0.0249998	integral membrane protein
PRKCB	protein kinase C, beta 1	NM_008855.1	870019	6679344	11.34289	10.802098	0.5407967	1.4547757	-1.454775692	0.026091	protein amino acid phosphorylation, cellular calcium ion homeostasis, protein kinase activity, calcium channel regulator activity, metal ion binding, protein binding
LOC668706	similar to Ribosomal protein L12	XM_001003212.2	7100750	149262009	11.14489	10.745167	0.3997191	1.3192511	-1.319251063	0.0263707	N/A
SPCS1	signal peptidase complex subunit 1 homolog (S. cerevisiae)	NM_026911	730373	142386308	11.19601	10.324639	0.8733683	1.8319349	-1.831934921	0.0268114	signal peptide processing
LOC270186 (Gm1656)	gene model 1656	XM_194481.2	1.01E+08	38090035	8.781605	8.3144807	0.4673248	1.3825434	-1.382543385	0.0268259	N/A
TCEA2	transcription elongation factor A (SII), 2	NM_009326	6370546	133892334	11.00629	11.430493	-0.424198	0.7452528	1.341826503	0.0282421	DNA dependant transcription regulation
D11BWG0434E (Rnasek)	ribonuclease, RNase K	NM_173742			10.69966	10.248301	0.4512651	1.3672292	-1.367229165	0.0286744	carboxypeptidase activity, proteolysis
AI480556 (Scaf1)	SR-related CTD-associated factor 1	NM_001008422			10.57381	11.061126	-0.487318	0.7133501	1.401836189	0.0289788	RNA splicing and transcription
A230057G18RIK (Rnoc2)	retinal noncoding RNA 2	AI-028326			10.34153	10.941221	-0.599694	0.6598939	1.515395165	0.0292497	retinal rod cell development
GPR88	G-protein coupled receptor 88	NM_022427	2450435	11967952	9.342387	8.9322818	0.4101056	1.328783	-1.328783042	0.0304929	g-protein coupled receptor activity, signal transduction, inflammatory response
RPS3	ribosomal protein S3	NM_012052	1570017	115270974	12.6656	12.091264	0.5743401	1.4889962	-1.488996236	0.0317166	structural constituent of ribosome, translation, ribonucleoprotein complex
BC021381	cDNA sequence BC021381	NM_145382	6550180	110626028	9.317729	9.7263248	-0.408596	0.7633664	1.32739298	0.0317608	N/A
SPNB3	spectrin beta 3	NM_021287	6660088	55926126	13.26362	13.683903	-0.414286	0.7603908	1.332639009	0.031848	actin binding
EIF5A	eukaryotic translation initiation factor 5A	NM_181582	1500059	142380083	9.906386	9.5134084	0.3929775	1.3131007	-1.313100674	0.0335226	translational initiation, apoptosis
ARPC1A	actin related protein 2/3 complex, subunit 1A	NM_019767	70672	118130576	10.8661	10.374955	0.4912427	1.4056652	-1.405655195	0.0338156	actin binding
LY6C1	lymphocyte antigen 6 complex, locus C1	NM_010741			10.2972	9.763329	0.5338668	1.4478045	-1.447804537	0.0346498	defence response
EG620119	predicted gene, EG620119	XM_907240			13.09807	12.432060	0.6660082	1.586677	1.586677	0.0349133	N/A
SEPW1	selenoprotein W, muscle 1	NM_009156	3800079	110735412	13.11293	13.717182	-0.604252	0.6578124	1.520190192	0.0352198	cell redox homeostasis, selinium binding
D16WSU169E	DNA segment, Chr 15, Wayne State University 169, expressed	NM_198420	2360022	38259213	12.05002	12.436341	-0.386326	0.7650756	1.307060321	0.0361612	protein binding, signal transduction
XBP1	X-box binding protein 1	NM_013842	3840594	13775155	12.26332	11.851544	0.4317724	1.3468897	-1.346889676	0.0364738	DNA dependant transcription regulation
MT3	metallothionein 3	NM_013603	1450537	7305286	13.09153	13.887177	-0.79565	0.5760835	1.735859587	0.036848	metal ion binding, negative regulation of neurogenesis, cellular metal ion homeostasis, synaptic vesicle, electron carrier activity
LGMN	legumain	NM_011175	3610301	118130389	10.25879	9.7630017	0.4957877	1.4100904	-1.410090424	0.0376398	negative regulation of growth
Sep-15	selenoprotein	NM_053102	1780451	110825968	9.713269	9.3197804	0.3934885	1.3135658	-1.313565838	0.0380667	N/A
SLC6A8	solute carrier family 6 (neurotransmitter transporter, creatine), member 8	NM_133987	3940341	19527207	12.21682	12.628909	-0.412089	0.7615342	1.330611352	0.0384566	neurotransmitter transport
PDIA3	protein disulfide isomerase associated 3	NM_007952	2120270	112293263	11.63454	11.123606	0.5109355	1.4249739	-1.424973881	0.0396007	isomerase activity, cell redox homeostasis, positive regulation of apoptosis, extracellular space and endoplasmic reticulum
RHOBTB3	Rho-related BTB domain containing 3	NM_028493	4010687	133892620	10.19687	10.702304	-0.50543	0.7044503	1.419546601	0.0398906	protein binding
1810015C11RIK	RIKEN cDNA 1810015C11 gene	AI-007507	1.03E+08		9.692754	9.987906	-0.395152	0.7604093	1.315081266	0.0401597	N/A
HBB-B1	hemoglobin, beta adult major chain	AI-002394			13.68647	12.2972	1.2892698	2.4440432	-2.444043202	0.0430769	iron ion binding, oxygen transporter activity, hemopoiesis
HP1BP3	heterochromatin protein 1, binding protein 3	NM_001122897	2650575	33468902	13.13461	13.564651	-0.430044	0.7422389	1.347275102	0.0442537	DNA binding, nucleosome assembly
TRF	transferrin	NM_133977	1050193	118129942	11.74369	10.984026	0.759666	1.6930986	-1.693098555	0.0449078	iron ion transport
1500015O10RIK	RIKEN cDNA 1500015O10 gene	NM_024283	3130368	142360223	10.22315	9.5821819	0.640971	1.5593784	-1.559378366	0.0451011	extracellular localisation
SQSTM1	sequestosome 1	NM_011018	6550056	118130186	12.57752	12.13629	0.4448674	1.3612079	-1.361207874	0.0462061	cell differentiation
ARF5	ADP-ribosylation factor 5	NM_007480			10.36876	9.834836	0.5339256	1.4478635	-1.447863485	0.0469999	small GTPase mediated signal transduction, protein transport
KCTD6	potassium channel tetramerisation domain containing 6	NM_027782	4070436	29789206	10.12315	9.6848702	0.4382826	1.3549904	-1.354990363	0.0482045	potassium ion transport
ZIC1	zinc finger protein of the cerebellum 1	NM_009573	670113	70778755	9.796103	9.0559537	0.7402493	1.6704645	-1.670464481	0.0489005	CNS development
VPS29	vacuolar protein sorting 29 (S. pombe)	NM_019780	580687	9790284	10.27364	9.825442	0.448201	1.3643379	-1.364337932	0.0490627	golgi to vacuole transport
LOC268569	hypothetical gene supported by NM_016750	XM_195821.2	1.04E+08	38050556	11.71762	10.917879	0.7997381	1.7407851	-1.740785126	0.0490876	N/A
	kinase 3										activity; nucleotide binding; magnesium ion binding
LOC665386	similar to prothymosin, alpha (gene sequence 28)	XR_032236.1	1.02E+08	149263919	9.819605	9.4166233	0.4041813	1.3233377	-1.323337706	0.0124395	N/A
ATP2B1	ATPase, Ca transporting, plasma membrane 1	NM_026482	1.04E+08	62234486	9.977962	9.4683417	0.5096201	1.4236752	-1.423675215	0.0126566	calcium ion transport
TMEM63B	transmembrane protein 63b	NM_198167	730397	142376394	10.41436	10.831194	-0.416836	0.7490654	1.334996983	0.0130026	
AARS	alanyl-tRNA synthetase	NM_146217	1570279	40254406	12.12966	12.51227	-0.382614	0.7670466	1.303701801	0.0134896	ATP binding, ligase activity, alanyl tRNA aminoacylation
CD9	CD9 antigen	NM_007657	4730041	47271531	10.27534	9.652769	0.622573	1.5396195	-1.539618547	0.015107	integral to membrane, negative regulation of cell proliferation, paranodal junction assembly, cell adhesion
5730589K01RIK (Tox4)	TOX high mobility group box family member 4	NM_023434			9.724976	9.3256281	0.3993474	1.3189112	-1.318911212	0.0153211	DNA dependant transcription regulation
TMEM25	transmembrane protein 25	NM_027865	2470575	142373326	11.30107	11.712962	-0.411908	0.7516287	1.330444229	0.0160309	integral membrane protein, molecular function
A930034L06RIK	RIKEN cDNA A930034L06 gene	NM_175692	4010332	124286808	12.12258	12.957388	-0.834812	0.560656	1.783624829	0.0163639	N/A
EIF4A2	eukaryotic translation initiation factor 4A2	NM_013506	1170494	7305018	10.03394	9.3589568	0.6750874	1.5966936	-1.596693551	0.0175745	ATP binding, helicase activity, nucleotide binding, translation initiation factor activity
HPCAL4	hippocalcin-like 4	NM_174998	2600484	146198535	13.16569	13.615098	-0.459409	0.7272841	1.374978461	0.0179233	calcium ion binding
LOC639086	similar to NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 11, mitochondrial	XR_034245.1	1.06E+08	149254444	10.88029	10.451041	0.429247	1.3465306	-1.346530621	0.0190052	N/A

Table 4.6.3: Dysregulated genes in Bupropion treated Q31L adults. Table shows the genes that are dysregulated in Q31L adult mice treated with the anti-depressant Bupropion, compared to Q31L drug naïve adult mice. Each table shows probe target ID (gene symbol), full gene name, accession number, Illumina probe ID, GI number, log transformed expression data, expression difference, anti-log of expression difference (to correct for negative values), fold-change, p-value and GeneOntology terms where available.

The potential importance of genes corrected by drug treatment was assessed by where they fell in the non-drug treated groups. For the L100P rolipram treated mice, *Bex2* was sited in the 3rd network produced with a good network score and high probability that the grouping was not by chance. It also, interestingly, made indirect connections with NFKβ which has been implicated in major mental illness and forms a central connection in this network for many other genes. *Pcbp1, Hnrpk* and *Wbp4* all appear in network 4 alongside *Egr2, Egr3* and *Egr4* which have been previously linked to major mental illness[238]. *Egr2* and *Wbp4* have a direct connection while *Pcbp1* and *Hnrpk* connect to each other and an Actin complex (Figure 4.6)

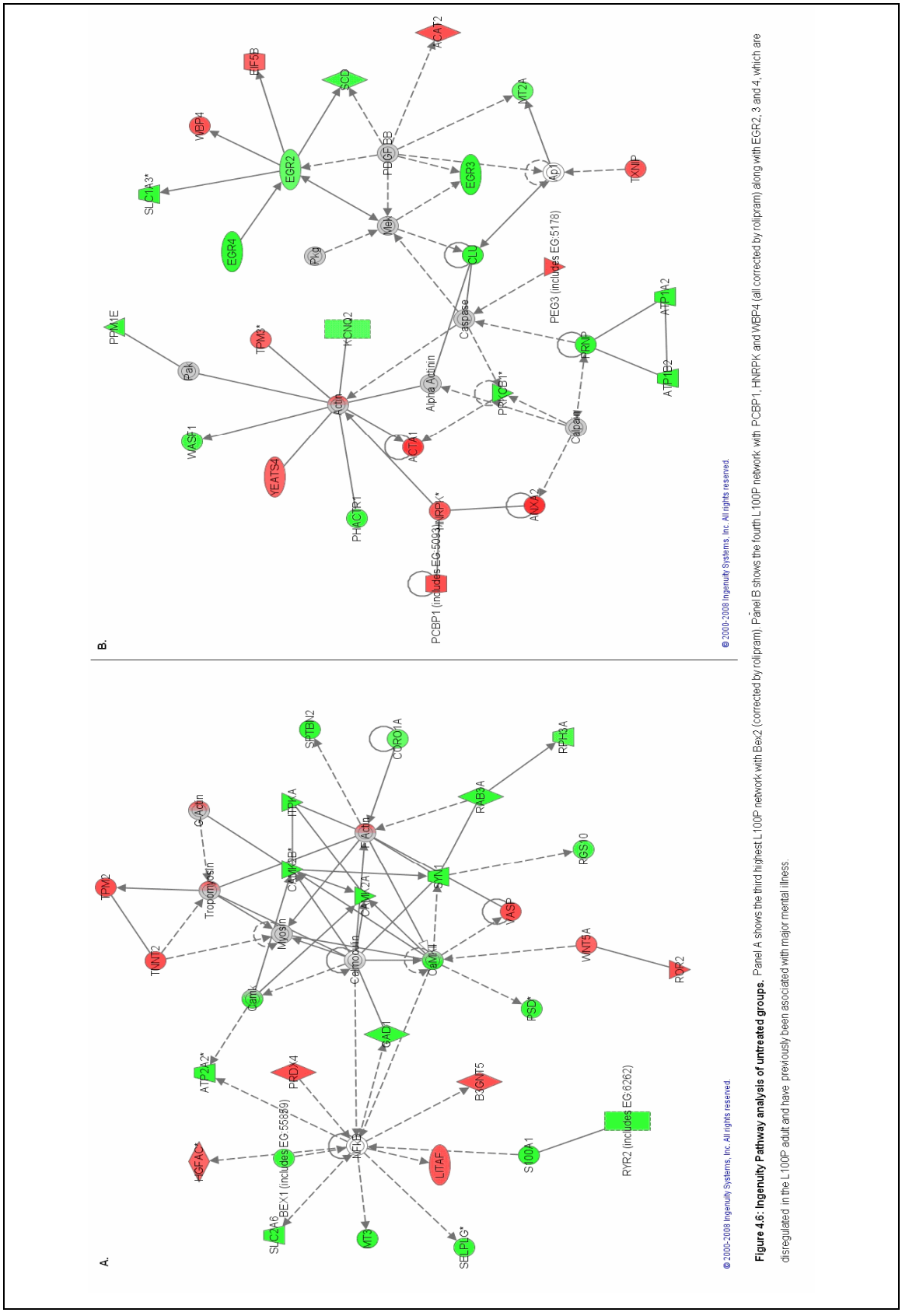

Figure 4.6: Ingenuity Pathway analysis of untreated groups. Panel A shows the third highest L100P network with Bex2 (corrected by rollpram), Panel B shows the fourth L100P network with PCBP1, HNRPK and WBP4 (all corrected by rollpram) along with EGR2, 3 and 4, which are disregulated in the L100P adult and have previously been associated with major mental illness.

4.3 Selection of Genes of Interest for Validation by QT RT PCR

The analysis outlined above resulted in a list of genes that were dysregulated in the Disc1 mutant mice compared to C57BL/6J wild-type controls, and genes which showed correction of expression in drug treated mutants. This list totalled 835 genes with some overlap between groups. Obviously it was not going to be possible to validate all of the genes highlighted so it was necessary to determine which genes most warranted follow up. This was done by scoring genes based on fold change, p-value, IPA analysis, over-enrichment of GO terms, presence in other groups from the array and overlaps with previously published studies on major mental illness. Genes were ranked within each pairwise comparison using a simple tally system, where each of the above selection criteria would add one tally to the gene 'score', and then the top 10% of genes were carried forward for validation. The list of genes taken forward for validation are shown in table 4.7.

TargetID	Full Gene Name	Comparisson	ProbeID	GI numbers	fold change	Pval	GO Terms	Evidence
DLG2	discs, large homolog 2 (Drosophila)	L100P/C57BL6	4230095	118136296	-1.55124522	0.019313055	protein binding, sensory perception of pain, synaptic transmission	L100P Adult microarray (-1.55FC) Walsh/Xu and pathway
NRGN	neurogranin	L100P/C57BL6	4280433	145587078	-2.823430807	0.021773197	protein kinase cascade, calmodulin binding	L100P Adult microarray (-2.8FC)
PLK1	polo-like kinase 1 (Drosophila)	L100P/C57BL6	1780369	128485537	1.640695258	0.035377774	protein amino acid phosphorylation, transferase activity, protein kinase activity, ATP binding	L100P Adult microarray (1.6) T1:11, network2
SYN1	synapsin I	L100P/C57BL6	4920301	62177179	-2.004335459	0.028394057	neurotransmitter secretion	L100P Microarray (-2.0), network3
CAMK2A	calcium/calmodulin-dependent protein kinase II alpha	L100P/C57BL6	1740333	118130259	-2.360244517	0.022981063	regulation of neurotransmitter secretion, protein amino acid phosphorylation, calcium ion transport, calcium and calmodulin dependant protein kinase complex, transferase activity, synapse	L100P Array (-2.3) IPA network 3
GRB10	growth factor receptor bound protein 10	L100P/C57BL6	6980082	118129993	1.740636092	0.043868465	signal transduction, receptor activity, SH3/SH2 adaptor activity, MAPKKK cascade	L100P array (1.7), IPA network 1, Akt associates
UBL3	ubiquitin-like 3	L100P/C57BL6	6450458	134053909	-1.457605333	0.023371204	protein modification	L100P Adult microarray (-1.45FC) T1:11
SORT1	sortilin 1	L100P/C57BL6	1850242	34610210	-1.369270568	0.032413703	vesicle organisation and biogenesis, myotube differentiation, cell differentiation, endocytosis	L100P Adult microarray (1.36) F22
PLXNB2	plexin B2	L100P/C57BL6	106020176	149267067	1.328022332	0.04777319	positive regulation of axonogenesis	L100P Adult microarray (1.3) F22
LITAF	LPS-induced TN factor	L100P/C57BL6	6940671	9910577	1.33284265	0.026259808	protein binding, apoptosis, regulation of transcription	L100P Adult microarray (1.3) T1:11
B3GNT5	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 5	L100P/C57BL6	4780528	31542176	1.361101925	0.037421285	protein amino acid glycosylation, transferase activity	L100P Adult microarray (1.4) T1:11
EMID2	EMI domain containing 2	L100P/C57BL6	360400	19263337	2.121568814	0.042783163	phosphate transport, proteinaceous extracellular matrix	L100P Adult microarray (2.1), Walsh
GAD1	glutamic acid decarboxylase 1	L100P/C57BL6	2360035	145301579	-1.611035765	0.01542056	synaptic transmission, neurotransmitter biosynthetic process, carboxylic acid metabolic process, pyridoxal phosphate binding	L100P array (-1.6), IPA network 3, join Nfkb and calmodulin
SYP	synaptophysin	L100P/C57BL6	2690168	6678194	-2.466941147	0.025737744	synaptic transmission, endocytosis	L100P Array (-2.5) IPA network 1
APOE	apolipoprotein E	L100P/C57BL6	4200671	31981894	-2.089886207	0.037039825	lipid metabolism, lipid transport, regulation of gene expression	L100P array (-2.1) IPA network 1
SHANK3	SH3/ankyrin domain gene 3	L100P/C57BL6	4070348	10946785	-1.326602403	0.025465175	MAPKKK cascade, intracellular signalling, post synaptic membrane	L100P array (-1.3), in DLG2 network with dlgap3 and T1:11 genes
MFGE8	milk fat globule-EGF factor 8 protein	L100P/C57BL6	3440373	113865978	-1.469803467	0.012169796	positive regulation of phagocytosis, cell adhesion, external side of plasma membrane	L100P Adult microarray (-1.5FC) F22
GSTM1	glutathione S-transferase, mu 1	L100P/C57BL6	1940332	141803498	-1.438654613	0.021084051	metabolic process, transferase activity	L100P Adult microarray (-1.4FC)
ID1	inhibitor of DNA binding 1	L100P/C57BL6	360398	118130032	1.305241528	0.022284642	regulation of transcription, regulation of angiogenesis, multicellular organismal development	L100P Adult microarray (1.3) T1:11
TGFBI	transforming growth factor, beta induced	L100P/C57BL6	2060446	145966862	1.543792423	0.022706007	protein binding, cell adhesion, extracellular	L100P Adult microarray (1.5) T1:11
AKAP9	A kinase (PRKA) anchor protein (yotiao) 9	L100P/C57BL6	510184	125661047	1.302202632	0.016816231	kinase activity, protein binding	L100P Adult microarray (1.3), Camargo
MLL5	myeloid/lymphoid or mixed-lineage leukemia 5	L100P/C57BL6	104540338	51710848	1.447917386	0.000392104	metal ion binding, regulation of transcription	L100P array (1.4), Walsh CNV
SLC1A3	solute carrier family 1 (glial high affinity glutamate transporter), member 3	L100P/C57BL6	1500070	40254195	-1.628570632	0.034156635	dicarboxylic acid transport, cell projection, transmembrane transport	L100P array (-1.6), Walsh structural variant
GP1BB	glycoprotein Ib, beta polypeptide	L100P/C57BL6	460725	50345975	-1.47975208	0.003219976	hemostasis, cell adhesion, protein binding	L100P array (-1.5), Xu CNV
ARMC10	armadillo repeat containing 10	L100P/C57BL6	5290575	146135072	1.439293987	0.027901153	regulation of cell growth	L100P array (1.4), Walsh CNV
PPM1E	protein phosphatase 1E (PP2C domain containing)	L100P/C57BL6	102480465	118130345	-1.492726007	0.040437896	protein amino acid dephosphorylation, metal ion binding, hydrolase activity, catalytic activity	L100P array (-1.5), Camargo interactome
GNB1	guanine nucleotide binding protein (G protein), beta 1	L100P/C57BL6	2120397	111186467	1.472410303	0.019239266	sensory perception of taste, Ras protein signal transduction, G-protein coupled receptor signalling pathway, muscarinic pathway, cell proliferation	L100P array (1.5), Camargo interactome
LOC100046895 (QKI)	similar to Quaking protein	L100P/C57BL6	105670403	149268446	1.525681046	0.001008059	N/A	L100P array (1.5), IPA network 1
DLGAP3	discs, large (Drosophila) homolog-associated protein 3	L100P/C57BL6	4540113	142351843	-1.339285878	0.041540165	cell-cell signalling, synapse, post synaptic membrane	L100P array (-1.3), in DLG2 network with shank3 and T1:11 genes
1500015O10RIK	RIKEN cDNA 1500015O10 gene	Q31L/C57BL6	3130368	142360223	1.460368798	0.031717451	extracellular localisation	Q31L Microarray (1.4) Q31L Bupro (-
ERDR1	erythroid differentiation regulator 1	Q31L/C57BL6	5890184	19111169	-1.574492787	0.007874278	biological process unknown, cellular component	Q31L Microarray
NDN	necdin	Q31L/C57BL6	5670075	116642896	1.359457738	0.002316554	transcription, sensory perception of pain, neuron migration, nerve growth factor receptor signalling pathway, axonogenesis, axonal extension, gamma tubulin binding, protein binding	Q31L Microarray
BEX2	brain expressed X-linked 2	L100P/L100PRoli	5670433	94408188	-1.485337179	0.00010109	nucleus and cytoplasm, no further info	L100P Adult (-1.5) and L100P Rolipram (+1.4). Walsh and Xu
PCBP1	poly(rC) binding protein 1	L100P/L100PRoli	1050088	47271536	-1.750607779	0.040927311	mRNA processing, translation activator activity, ribonucleoprotein complex	L100P Microarray (1.43) L100P Rolipram (-1.75)
HNRPK	heterogeneous nuclear ribonucleoprotein K	L100P/L100PRoli	1190348	142350515	1.337295293	0.042310367	mRNA processing, RNA splicing	L100P Microarray (1.3) L100P Rolipram (-1.5)
WBP4	WW domain binding protein 4	L100P/L100PRoli	2230019	124286863	1.354214316	0.004629468	mRNA processing and RNA splicing	L100P Microarray (1.35) L100P Rolipram (-1.34)
HES6	hairy and enhancer of split 6 (Drosophila)	L100P/L100PRoli	540411	142369243	-1.389951192	0.043066239	negative regulation of neuron differentiation, nervous system development	L100P Microarray (1.67) L100P Rolipram (-1.38)
D130059O18RIK (Sfrs6)	splicing factor, arginine/serine-rich 6	L100P/L100PRoli	106100411		-1.393737236	0.025635345	RNA splicing, mRNA processing	L100P Microarray (1.3) L100P Rolipram (-1.39)
CORO1B	coronin, actin binding protein 1B	L100PE/C57BL6E Q31LE/C57BL6E	6350592	142388278	-1.724758876	0.012092161	actin binding, cytoskeleton	L100P Embryo (-1.6) and Q31L Embryo (-1.7) microarray
PDE4D	phosphodiesterase 4D, cAMP specific	L100PE/C57BL6E Q31LE/C57BL6E	2470528	118130222	1.938462264	0.00865751	cAMP catabolism and signal transduction	L100P Embryo (1.94) and Q31L Embryo (1.4) microarray
DMRT2	doublesex and mab-3 related transcription factor 2	L100PE/C57BL6E	2630068	146149265	-2.795713453	0.006649191	regulation of transcription, sex differentiation	L100P Embryo microarray (-2.8)
NUMB	numb gene homolog (Drosophila)	L100PE/C57BL6E Q31LE/C57BL6E	2450735	6754911	-1.646325327	0.039977665	neuroblast proliferation, axonogenesis	L100P Embryo (-1.65) and Q31L Embryo (-1.7) microarray
5830454D03RIK (Hook3)	hook homolog 3	Q31LE/C57BL6E	101990068		-1.301582321	0.026481508	microtubule cytoskeletal organisation and biogenesis	Q31L Embryo microarray (-1.3), T1:11

Table 4.7: Genes for validation by qRT PCR. Columns 1 and 2 show the gene symbol and full name, column 3 shows the comparison in which it is dysregulated, columns 4 and 5 show probe ID and GI numbers for the gene, column 7 and 8 show fold-change and p-value, column 9 gives GOterms and column 10 gives further evidence to support validation.

4.4 Identification of L100P outlier

During the Taqman qRT-PCR, a conversation with a more experienced microarray statistician lead me to take a closer look at the normalised raw data for the L100P mutant mouse. As the number of genes classified as differentially expressed were very high it was thought that there may be an anomaly in the data skewing the result (Dr Freeman, personal communication).

I picked twenty genes at random through the microarray dataset and ran interquartile-range outlier analysis on the four L100P adult (saline) pools (3-4 mice per pool) which confirmed one pool, L100P Sal F2, as an outlier in 15 out of the 20 genes analysed. As there were 46,644 genes in the total dataset, it was not possible to test every gene to see whether this pool was an outlier therefore this pool was removed and the entire dataset reanalysed.

The qRT PCR validation that had been run so far (22 genes on each mouse sample) suggested that only one mouse in the pool was an outlier by inter-quartile range analysis. Of the outliers removed during analysis of the Taqman qRT-PCR data, 40% were mice from the L100P Sal F2 pool. Given that there were four L100P adult (saline) pools this is higher by ~15% than expected than random chance of 25%. 87.5% of the time the L100PSalF2 pool sample removed was B100-1-32 (14 times in 22 analyses). As the array data was run on the complete pool, all mice in the pool had to be removed from further analysis. The genes already validated by Taqman qRT-PCR would be reanalysed and will be described in chapter 5.

After identifying this pool I went back and looked at the behavioural data from the PrePulse Inhibition trial to determine if the mice in this pool had displayed behaviours which would be suggestive of different genetic expression (figure 4.7).

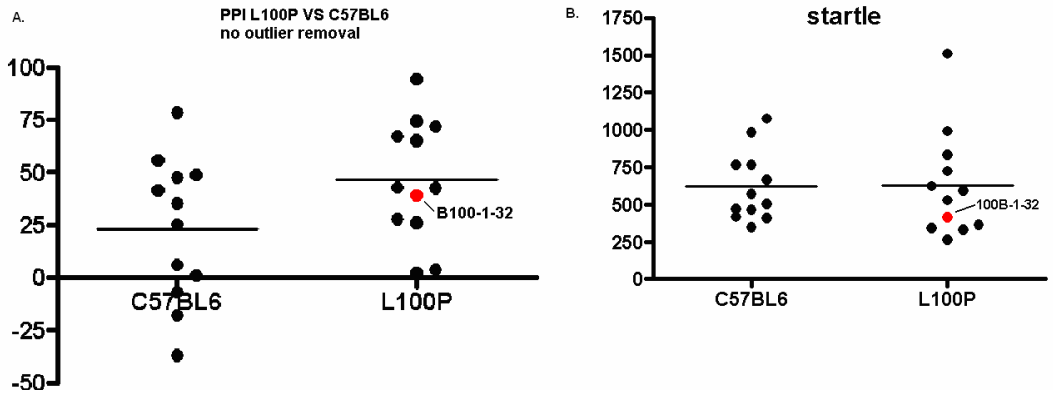

Figure 4.7: PrePulse Inhibition data for L100PSaline and C57BL/6J mice. Panel A shows the total PPI values with no outlier removal. Panel B shows the total startle response with no outlier removal. Sample B100-1-32 is the sample most removed as an outlier in the genetic validation stage, shown here as not being an outlier in the behavioural study.

As can be seen from the graph in figure 4.7, the B100-1-32 mouse did not behave differently to the other L100Psaline mice tested despite its aberrant (outlier) gene expression.

4.4.1 Reanalysis of array dataset

The identification of the outlier group necessitated complete re-analysis of the entire array dataset post the normalisation step (see chapter 3 section 7.2). Removal of the L100Psaline F2 pool left seven samples in the L100P saline group, which still gave 80% power to detect a 1.3 fold-change. Analysis was carried out as described initially in chapter 3, and chapter 4 section 2.

Removal of the L100Psaline pool altered the detection p-values for each probe and this gave rise to a completely new dataset for analysis. Of the 46,644 successful probes, 20,302 (43.5%) probes were expressed above background. This is slightly higher than expected [231]. As described in Chapter 3, the next filtering step was to remove all probes with less than 1.3 fold-change in the dataset. This reduced the number of probes in the L100Psaline set to 1272, in the Q31Lsaline group to 26, in the L100PEmbryo to

1344, in the Q31LEmbryo to 613, in the L100PRolipram to 456, in the L100PClozapine to 321 and in the Q31LBupropion to 262 (Table 4.8)

Finally, using the same methods as described in section 4.2, the probes that survived the fold-change filter were tested for differential expression using the Analysis of Variance model (ANOVA) (table 4.8).

Group	Control	Number of Probes Expressed above Background	Number of Probes Surviving fold-change filter	Number of Probes Differentially Expressed Compared to Controls (% dataset)
L100P Saline	C57BL6 Saline	18067	1272	88 (0.487)
Q31L Saline	C57BL6 Saline	17700	26	19 (0.107)
L100P Embryo	C57BL6 Embryo	18078	1344	47 (0.260)
Q31L Embryo	C57BL6 Embryo	18097	613	25 (0.138)
L100P Clozapine	L100P Saline	17491	321	4 (0.023)
L100P Rolipram	L100P Saline	17830	456	36 (0.202)
Q31L Bupropion	Q31L Saline	16871	262	186 (1.102)

Table 4.8 Summary of probes differentially expressed by genotype of drug treatment. Column 1 shows the group of interest and column 2 the control group for the comparison. Column 3 shows the number of genes expressed above background, column 4 shows the number of probes surviving the fold-change filter, and column 5 shows the number of probes, and percentage of those probes that were significantly differentially expressed (Fs p-value ≤ 0.05).

In total 368 genes were found to be differentially expressed in this dataset (table 4.9). 66 of these genes were also present in the previous dataset, and overlap of only 18%.

Gene_ID	Probe ID	Comparisson	Fold change	Pval	Function (and alias)	Localisation	Previous Implications in Major Mental Illness: Disc1 Interaction	Overenrichment GO-Tree	IPA Top Canonical Pathways
D930049F02RIK	1064000020	L10P/C57BL6	-1.556407196	8.05E-10	protein binding	-	-	-	-
REC8	5900014	L10P/C57BL6	-1.570964298	1.48E-08	synapsis	synaptotemal complex	-	-	-
ENTPD4	7040288	L10P/C57BL6	-1.594896685	0.00000005	calcium ion binding	cytoplasmic vesicle, golgi apparatus	-	yes	-
SLC35F4	1069401B0	L10P/C57BL6	-1.271976785	9.63E-08	protein binding, transport	membrane	-	-	-
S100A9	70112	L10P/C57BL6	1.612692265	0.000000163	calcium ion binding, chemotaxis	-	-	-	yes
MSR2	60204	L10P/C57BL6	-1.25900061	0.000000202	receptor activity (immune)	-	-	-	-
DBT	5220364	L10P/C57BL6	-1.327292297	0.000000629	metabolic	mitochondrion	-	yes	yes
ABHD4	6760450	L10P/C57BL6	-1.592240154	0.000001108	lipid catabolic process, proteolysis	-	-	-	-
CHAC1	7100725	L10P/C57BL6	-1.304238931	0.000003396	protein binding	cellular_component	-	-	-
EGR3	6940128	L10P/C57BL6	-1.260215363	0.0000109	neuromuscular synaptic transmission	nucleus	Yamada et al (2006)	yes	yes
SCL00012841_18	1064001O0	L10P/C57BL6	1.308686672	0.0000116	-	-	-	-	-
CORT	360008	L10P/C57BL6	-1.25000147	0.0000167	hormone activity	extracellular region	-	-	-
S100A9	7050528	L10P/C57BL6	1.277056227	0.0000235	calcium ion binding, leukocyte chemotaxis	-	-	-	yes
BEX2	5670433	L10P/C57BL6	-1.290091214	0.0000381	apoptosis/cell cycle	cytoplasm/nucleus	-	-	yes
SLC4A7	2370056	L10P/C57BL6	1.354516275	0.0000383	sodium ion transport	plasma membrane	-	-	-
PSMD1	1024701S6	L10P/C57BL6	1.270656082	0.0000441	regulation of protein catabolic process	proteasome complex	-	yes	-
RAPGEF5	1500603	L10P/C57BL6	1.385244486	0.0000457	regulation of small GTPase mediated signal transduction	nucleus	-	yes	-
LOC100047638		L10P/C57BL6	1.424550103	0.0000603	-	-	-	-	-
DCN	510332	L10P/C57BL6	1.329668414	0.0000676	protein binding	proteinaceous extracellular matrix	-	-	yes
SUSD4	1850017	L10P/C57BL6	1.515679987	0.0000901	-	integral to membrane	-	-	-
EXOC4	2470263	L10P/C57BL6	1.256813183	0.0000925	vesicle docking during exocytosis	exocyst	-	yes	yes
DNAJB5	2900215	L10P/C57BL6	-1.295363147	0.000098	heat shock protein binding	-	-	yes	-
D14ERTD449E	1340440	L10P/C57BL6	-1.285292956	0.0000983	-	integral to membrane	-	-	-
PTPRT	8620433	L10P/C57BL6	1.25279485	0.0000993	protein tyrosine phosphatase activity	membrane	-	-	-
IMPAD1	2690594	L10P/C57BL6	1.278967778	0.0001035	hydrolase activity	-	-	-	-
SCL00011181_0	1069401O6	L10P/C57BL6	1.384539772	0.0001194319	-	-	-	-	-
PYGO2	360494	L10P/C57BL6	-1.262483422	0.00036828	brain development, Wnt receptor signalling pathway	nucleus	-	yes	-
LOC331507	1047601O9	L10P/C57BL6	1.499064967	0.0003641123	alias Gm1866 - function unknown	-	-	-	-
BOK	1170373	L10P/C57BL6	-1.313393899	0.0003659325	regulation of apoptosis	intracellular	-	-	-
A330057G13RIK	1032900O1	L10P/C57BL6	1.249999672	0.0003775383	-	-	-	-	-
BC003331	1980164	L10P/C57BL6	1.251822372	0.0003786	oxidoreductase activity	membrane	-	yes	-
GRIA1	1340152	L10P/C57BL6	1.304983423	0.0004923302	receptor internalisation, ion transport	postsynaptic density	-	yes	yes
LOC100047606	6100484	L10P/C57BL6	1.445129869	0.0005091744	Alias NTRK3 - kinase	plasma membrane	-	-	yes
NAPB	1001105524	L10P/C57BL6	1.294447461	0.0006324624	syntaxin binding	membrane	-	yes	yes
HBG1	2900619	L10P/C57BL6	1.269847831	0.0007101411	oxygen transporter activity	-	-	yes	-
PJA2	430537	L10P/C57BL6	1.539366912	0.0008071597	ubiquitin dependent protein catabolic process	synapse, cell junction	-	-	-
NRXN1	460408	L10P/C57BL6	1.352061316	0.00113402	synaptic transmission	postsynaptic membrane	Kirov et al 2008, Need et al 2009, Vrijenhoek et al 2008, Rujescu et al 2009	yes	-
PPARGC1A	4670040	L10P/C57BL6	1.284296082	0.001167898	positive regulation of transcription from RNA polymerase II promoter	nucleus	-	yes	yes
C13002E19RIK	1014105A6	L10P/C57BL6	1.304679041	0.001199618	-	-	-	-	-
MTAP1B	2230332	L10P/C57BL6	1.461508929	0.001468155	Alias MAP1B - microtubule-based process	microtubules, cytosol	-	yes	yes
D11BWG0617E	1009401S4	L10P/C57BL6	1.260301483	0.00160369	nucleotide binding	nucleus	-	-	-
1500012F01RIK	1048601O4	L10P/C57BL6	-1.260386824	0.001736065	transcription regulator activity	-	-	-	-
PSMA4	4560427	L10P/C57BL6	1.25218406	0.001740174	threonine-type endopeptidase activity	proteosome core complex	-	-	-
GNAI1	4560390	L10P/C57BL6	1.295630328	0.00177657	G-protein coupled receptor protein signaling pathway, signal transduction	-	-	-	yes
A330021D07RIK	1041201A2	L10P/C57BL6	1.287174767	0.001823364	-	-	-	-	-
PDE1A	1050458	L10P/C57BL6	1.267445993	0.002243739	calcium- and calmodulin-regulated 3',5'-cyclic-GMP phosphodiesterase activity	-	-	yes	-
MLL5	1045403S8	L10P/C57BL6	1.289732863	0.002456157	chromatin modification, cell cycle arrest	nucleus	Walsh et al 2008	yes	-
FBXO41	1032900O4	L10P/C57BL6	1.333240204	0.002924403	ubiquitin dependent protein catabolic process	intracellular	Camargo et al 2006	-	-
DBP	4200270	L10P/C57BL6	1.278119282	0.002981867	transcription regulator activity	nucleus	-	yes	yes
C730006J16	1067604B8	L10P/C57BL6	1.268483689	0.003052346	-	-	-	-	-
SEC24D	3060279	L10P/C57BL6	-1.280226252	0.003299153	vesicle mediated transport	golgi apparatus	-	yes	-
F5	1990619	L10P/C57BL6	1.266897118	0.003382816	blood coagulation/circulation	extracellular region	Disregulated in t1:11 translocation family (unpublished)	-	-
SCL00009201_21	1017700B6	L10P/C57BL6	1.269161784	0.003917794	-	-	-	-	-
2900060B14RIK	1028100J9	L10P/C57BL6	1.814022861	0.004334075	-	-	-	-	-
LOC361445	1070003G2	L10P/C57BL6	1.290070697	0.006686057	-	-	-	-	-
ATP5L	1015000T0	L10P/C57BL6	1.273454521	0.007337984	proton-transporting ATP synthase complex, coupling factor F(o)	mitochondrion	-	yes	-
2900019G14RIK	1058060O4	L10P/C57BL6	1.449124131	0.009705176	alias Slc1a2 - high-affinity glutamate transmembrane transporter activity	axon	-	-	-
GRIA2	5080088	L10P/C57BL6	1.265791777	0.011056047	synaptic transmission	synapse	-	yes	yes
6430510M02RIK	4540270	L10P/C57BL6	1.299009315	0.012539956	alias Strip - RNA binding	microtubule cytoskeleton	-	-	-
CPLX2	6980538	L10P/C57BL6	1.41073936	0.013852808	neurotransmitter transport	dendrite	-	yes	-
ARC	4610093	L10P/C57BL6	-1.346512662	0.014205781	actin binding post synaptic membrane	cell projection/post synaptic membrane	-	yes	-

Table 4.9: New list of dysregulated genes post outlier removal. Continued next page…….

Gene_ID	Probe ID	Comparison	Fold change	Pval	Function (and alias)	Localisation	Previous Implications in Major Mental Illness: Disc1 Interaction	Overenrichment GO-Tree	IPA Top Canonical Pathways
THSD4	1990093	L100P/C57BL6	1.265217478	0.01464925	metalloendopeptidase activity	extracellular region	-	-	-
6330407J23RIK	1780004	L100P/C57BL6	-1.261736823	0.01463508	protein binding	membrane	-	-	-
NDUFB6	10327041	L100P/C57BL6	-1.260073284	0.01465679	electron transport chain	mitochondrial inner membrane	-	-	-
2900016B01RIK	10365064	L100P/C57BL6	1.401067801	0.01634849	alias Kbtbd11 - function unknown	-	-	-	-
DUSP6	5910286	L100P/C57BL6	-1.279670866	0.01756764	MAP kinase tyrosine/serine/threonine phosphatase activity	cytoplasm	Disregulated in t(1;11 translocation family (unpublished)	yes	yes
IQGAP2	1042003900	L100P/C57BL6	1.262757688	0.01910184	small GTPase mediated signal transduction	membrane	-	-	-
C330006P03RIK	1014503593	L100P/C57BL6	-1.452372072	0.01993585	-	-	-	-	-
CD69A	4610441	L100P/C57BL6	-1.261256736	0.02067036	protein binding	anchored to membrane	-	yes	-
FOLR1	2470369	L100P/C57BL6	1.336440864	0.02085806	posttranslational protein targeting to membrane	membrane	Novak et al 2009: disregulated in t(1;11 translocation (unpublished)	yes	-
B230375D24RIK		L100P/C57BL6	1.251059613	0.02176949	Alias Nrxn3 - calcium channel regulator activity	presynaptic membrane		-	-
2610040L17RIK	1001302073	L100P/C57BL6	1.267543605	0.02369493	alias Odz2 - signal transduction	membrane	-	-	-
SGTB	460170	L100P/C57BL6	1.295368468	0.02684823	protein homodimerization activity	-	-	yes	-
ZBTB7A	3780092	L100P/C57BL6	1.259928942	0.02944908	regulation of transcription, DNA-dependent, cell differentiation	nucleus	-	yes	-
DDX6	2630025	L100P/C57BL6	1.252047641	0.02948104	ATP binding	cytoplasm	-	yes	yes
0610007J10RIK	10567098	L100P/C57BL6	1.271057707	0.04612645	-	-	-	-	-
POP4	4560600	L100P/C57BL6	-1.6203888	0.00618728	rRNA processing/cleavage	ribonuclease MRP complex	-	-	-
		L100P/L100PRoli	1.83990718	0.05587922					
		Q31L/Q31LBugro	-1.712103583	0.04966937					
A930034L06RIK	4010332	L100P/C57BL6	1.486232383	0.0009034	alias Snglh11 - function unknown	-	-	-	-
		Q31L/Q31LBugro	1.78636248	0.01636394					
PAK3	4210136	L100P/C57BL6	1.3456638	0.00017985	kinase activity	-	Blemenu et al 2000, Donelly et al 1996, Peippo et al 2007	yes	yes
		L100P/L100PRoli	-1.28061126	0.04924822					
NAV/BP5-PENDING (Ndfip)	1003800053	L100P/C57BL6	-1.447086508	0.00082845	protein binding	mitochonrion	-	-	-
		L100P/L100PRoli	-1.72396641	0.01917756					
		Q31L/Q31LBugro	-1.86021361	0.02400219					
LOC674214	1570138	L100P/C57BL6	-1.23926874	0.00687462	-	cytoskeleton	-	-	-
		L100P/L100PRoli	-1.27268557	0.01756453					
2900040C04RIK	10467008	L100P/C57BL6	1.276396567	0.02334873	-	-	-	-	-
		Q31L/C57BL6	1.422670229	0.03026656					
WDFY1	1580463	L100P/C57BL6	1.458124652	2.85E-10	metal ion binding	-	-	yes	-
		Q31L/C57BL6	1.504868507	4.94346E-11					
POU6F1	6200168	L100P/C57BL6	-1.26621471	0.00017496	regulation of transcription, DNA-dependent	nucleus	-	-	-
		Q31L/Q31LBugro	1.33108192	0.00327969					
EGR2	3800403	L100P/C57BL6	-1.261317751	0.00114863	brain segmentation, motor axon guidance, schwann cell differentiation	nucleus	Yamada et al (2006)	yes	yes
		Q31L/Q31LBugro	-1.30621938	0.00226526					
SYN2	3990693	Q31L/Q31LBugro	1.290757468	0.04789610	neurotransmitter secretion	synaptic vesicle membrane	-	yes	-
EGR4	3120750	L100P/C57BL6	-1.388262908	0.00553847	regulation of transcription, DNA-dependent	nucleus	Yamada et al (2006)	yes	yes
		Q31L/Q31LBugro	-1.60899127	0.00945907					
ATP2B1	10451018	L100P/C57BL6	1.463274442	0.01756814	calcium ion transport	plasma membrane	-	yes	yes
		Q31L/Q31LBugro	-1.423675215	0.01285662					
DDH4S114	2350537	L100P/L100PCloz	1.261019023	0.01625480	regulation of transforming growth factor beta signalling pathway	-	-	-	-
MTDNA_ATP6	10636013 2	L100P/L100PCloz	-1.267509894	0.04985798	hydrogen ion transmembrane transporter activity	mitochondrial inner membrane	-	-	-
TMEM68	1770456	L100P/L100PCloz	-1.263872269	0.06645624	acetyl transferase activity	membrane	-	-	-
DUSP1	6960121	L100P/L100PRoli	1.522828927	0.00451223	cell cycle, protein amino acid dephosphorylation	-	-	-	-
ZYX	4730068	L100P/L100PRoli	1.266576484	0.00672136	cell adhesion	cytoskeleton	-	-	-
C330002I19RIK	430451	L100P/L100PRoli	1.27256515	0.00942128	regulation of transcription	membrane	-	-	-
LOC675440	70010	L100P/L100PRoli	-1.344691191	0.01418161	-	-	-	-	-
PTGDS	3610619	L100P/L100PRoli	-1.434819161	0.02264662	lipid biosynthetic programme, transport	intra and extracellular	-	-	-
LOC100044592	6620452	L100P/L100PRoli	-1.35516435	0.02372873	-	-	-	-	-
ZFP36	4010735	L100P/L100PRoli	-1.297734545	0.02635441	apoptosis, response to DNA damage	cytoplasm	-	-	-
2410129H14RIK	4010736	L100P/L100PRoli	-1.29972445	0.02639341	-	-	-	-	-
TIMP4	6110433	L100P/L100PRoli	-1.303105457	0.02626278	metalloendopeptidase activity	extracellular region	-	-	-
LOC381891		L100P/L100PRoli	-1.331523084	0.02899786	-	-	-	-	-
2810008M24RIK	10542047 1	L100P/L100PRoli	-1.279601591	0.03498773	RIK-EN mouse cDNA	membrane	-	-	-
SCL0004020.1_31	10313018 1	L100P/L100PRoli	-1.537052174	0.03967371	-	-	-	-	-
PRR18	2030064	L100P/L100PRoli	1.255013036	0.04069620	G-protein coupled receptor signalling pathway	endoplasmic reticulum	-	-	-
RANBP9	4670685	L100P/L100PRoli	-1.331496847	0.04420128	RAN GTPase binding	cytoplasm	-	-	-
5730437N04RIK	2320022	L100P/L100PRoli	-1.461298255	0.04425862	-	membrane	-	-	-
2310008I22RIK	10295070 6	L100P/L100PRoli	-1.264121511	0.04441158	alias March5 - ubiquitin dependent protein catabolic process	membrane	-	-	-
ATP5B	3870138	L100P/L100PRoli	-1.36877392	0.04752810	negative regulation of cell adhesion involved in substrate bound cell migration	membrane	-	-	-
LOC433749	580142	L100P/L100PRoli	-1.542010105	0.05040070	-	-	-	-	-
HNRPK	1190348	L100P/L100PRoli	-1.306706786	0.05041176	RNA splicing	splicosome	-	-	-
HNRPH1	1170086	L100P/L100PRoli	-1.323078399	0.05269548	mRNA processing	splicosome	-	-	-
NARFL	2260372	L100P/L100PRoli	1.27325620	0.01119670	-	nucleus	-	-	-
		Q31L/Q31LBugro	1.29869283	0.02271480					
PPP1CB	3130411	L100P/L100PRoli	-1.27566094	0.01808463	cell cycle, cell division	cytoplasm	-	-	-
		Q31L/Q31LBugro	-1.27566094	0.01808463					
SEPT15	1780451	L100P/L100PRoli	-1.36541674	0.02719257	de novo post translational protein folding	endoplasmic reticulum	-	-	-
		Q31L/Q31LBugro	-1.313966683	0.0380667					
TCEA2	6370546	L100P/L100PRoli	1.28953170	0.00356081	regulation of transcription	nucleus	-	-	-
		Q31L/Q31LBugro	1.341935503	0.02824212					
MAST3	450402	L100P/L100PRoli	1.26400011	0.03245639	protein amino acid dephosphorylation	-	-	-	-
		Q31L/Q31LBugro	1.339891194	0.01239628					
EG633120	1017406 47	L100P/L100PRoli	-1.31416534	0.03371972	predicted gene - transporter activity	membrane	-	-	-
		Q31L/Q31LBugro	-1.31416534	0.03371972					
SQSTM1	6560056	L100P/L100PRoli	-1.48263838	0.04793096	cell differentiation, apoptosis	cytosome, endosome	-	-	-
		Q31L/Q31LBugro	-1.36120787	0.04620608					
PDIA3	2120270	L100P/L100PRoli	-1.371567775	0.04978069	positive regulation of apoptosis	endoplasmic reticulum	-	-	-
		Q31L/Q31LBugro	-1.42297388	0.0395007					

Table 4.9: New list of dysregulated genes post outlier removal. Continued next page…….

Gene_ID	Probe ID	Comparison	Fold change	Pval	Function (and alias)	Localisation	Previous Implications in Major Mental Illness/ Disc1 Interaction	Overenrichment GO-Tree	IPA Top Canonical Pathways
XPOT	7050184	L100PE/C57BL6E	1.338517213	0.001879016	tRNA binding	nucleus	-	-	-
LOC100044475	2510010	L100PE/C57BL6E	-1.397978867	0.003778124	-	-	-	-	-
1700074P13RIK	5670161	L100PE/C57BL6E	1.573697354	0.005846003	peptidase activity	golgi apparatus	-	-	-
DDX46	1040100026	L100PE/C57BL6E	1.340680367	0.005946813	RNA splicing	nucleus	-	-	-
DMRT2	2630068	L100PE/C57BL6E	-2.796713453	0.006649191	transcription factor activity	nucleus	-	-	-
PRMT6	1053680154	L100PE/C57BL6E	-2.236587664	0.007013376	methyltransferase activity	nucleus	-	-	yes
OLFR1024	2407694	L100PE/C57BL6E	1.845721945	0.007960407	signal transduction	plasma membrane	-	-	-
LOC386917	1020680712	L100PE/C57BL6E	1.254620591	0.008007835	-	-	-	-	-
MYCN	110577	L100PE/C57BL6E	-1.266869363	0.009266669	positive regulation of cell proliferation	nucleus	-	-	yes
4833441J24RIK	3870142	L100PE/C57BL6E	1.51124216	0.011762463	alias THB - protein modification	-	-	-	-
SIT	60600	L100PE/C57BL6E	-1.605178383	0.012936778	-	membrane	-	-	yes
PPAPDC1	4560017	L100PE/C57BL6E	-1.286314412	0.014588752	catalytic activity	membrane	-	-	-
SCL0005681_2	1064550142	L100PE/C57BL6E	-1.2506784462	0.017939141	-	-	-	-	-
2810425O13RIK	1000550131	L100PE/C57BL6E	-1.368990246	0.020424339	alias Cab39l - protein binding	membrane	-	-	-
MTRR	4060358	L100PE/C57BL6E	-1.269890108	0.021260106	protein stabilisation	-	-	-	-
GPR103	5910066	L100PE/C57BL6E	-1.41840396	0.021428168	G-protein coupled receptor signal transduction	membrane	-	-	-
VTI1-PENDING	1024804046	L100PE/C57BL6E	1.485576832	0.021964229	-	membrane	-	-	-
2610028A01RIK	3360154	L100PE/C57BL6E	1.301669234	0.022719529	negative regulation of cell cycle	nucleus	-	-	-
LOC634350	1064301084	L100PE/C57BL6E	-1.301652614	0.023088732	-	-	-	-	-
A530020A01RIK	1019990112	L100PE/C57BL6E	-1.403045672	0.024389486	-	-	-	-	-
Sfxs12	1041220142	L100PE/C57BL6E	-1.311128875	0.025087404	splicing factor	nucleus	-	-	yes
GPR85	5130750	L100PE/C57BL6E	1.263868747	0.027784273	G-protein coupled receptor signal transduction	membrane	-	-	-
CDA	1960605	L100PE/C57BL6E	-1.304049524	0.028316119	cytidine metabolic process	-	-	-	yes
TIFP39	6200341	L100PE/C57BL6E	1.342726652	0.029857441	alias Phf2 - neuropeptide signalling pathway	extracellular region	-	-	-
VPS33B	7000279	L100PE/C57BL6E	1.291144297	0.029307052	vesicle docking during exocytosis	endosome	-	-	-
SLC35A3	4760066	L100PE/C57BL6E	-1.280063415	0.030194627	carbohydrate transport	golgi apparatus	-	-	yes
RHOX5	7040066	L100PE/C57BL6E	1.367191755	0.030742566	alias PEM - regulation of transcription	cytoplasm, nucleus	-	-	yes
LOC381443	1040700066	L100PE/C57BL6E	1.27841479	0.030957958	-	-	-	-	-
WDR32	450239	L100PE/C57BL6E	-1.259030705	0.034208135	alias DCaf10 - function unknown	-	-	-	-
KAZALD1	3800603	L100PE/C57BL6E	-1.497722673	0.036018152	cell differentiation and growth	extracellular region	-	-	-
LOC100048816	6301113	L100PE/C57BL6E	1.327310786	0.037810219	alias BAI2 - Gprotein coupled receptor	membrane	-	-	yes
8030474H12RIK	1066590100	L100PE/C57BL6E	1.316884304	0.038770871	-	-	-	-	-
E330009E22RIK	1017707003	L100PE/C57BL6E	-1.250360205	0.039114861	-	-	-	-	-
ATP4B	1740156	L100PE/C57BL6E	1.287915467	0.039466776	sodium ion transport	membrane	-	-	-
GSTM7	5570647	L100PE/C57BL6E	1.501862732	0.044843688	glucothionine transferase activity-metabolic	cytoplasm	-	-	-
LOC100044089	240739	L100PE/C57BL6E	-1.719825569	0.045568987	alias PIAS3 - transcription regulation	nucleus	-	-	yes
LOC380654	1060440451	L100PE/C57BL6E	-1.660357633	0.045767755	-	-	-	-	-
CSTF1	4610129	L100PE/C57BL6E	1.643501843	0.047501491	mRNA processing	nucleus	-	-	yes
CDH11	1230133	L100PE/C57BL6E	1.547422296	0.048900635	cell adhesion	membrane	-	-	yes
PDE4D	2470528	L100PE/C57BL6E	1.93846226	0.00965751O	cAMP catabolism, signal transduction	cytoplasm,cytoskeleton	Camargo et al 2006; Millar et al 2005	-	-
PRM3	4810577	Q31LE/C57BL6E	1.41679003	0.00691747	-	-	-	-	-
NUMB	2450735	L100PE/C57BL6E	-1.524309523	0.011445011	cell differentiation, mitotic chromosome condensation	nucleosome	-	-	-
B930001L07RIK	105360086	Q31LE/C57BL6E	1.307110238	0.03356575	-	-	-	-	-
D130004A15RIK	1020880155	L100PE/C57BL6E	-1.641963253	0.03997786	axonogenesis, forebrain development	membrane, nucleus	-	-	-
CORO1B	6350592	L100PE/C57BL6E	-1.706348305	0.022887492	-	-	-	-	-
COPE	2340068	Q31LE/C57BL6E	1.39067144	0.002348976	-	-	-	-	-
B230354K09RIK	1060510632	Q31LE/C57BL6E	1.269884255	0.015979337	-	-	-	-	-
6720463M19RIK	1061110204	Q31LE/C57BL6E	-1.775175999	0.01063187	-	-	-	-	-
5830454D03RIK	1019900868	Q31LE/C57BL6E	-1.301582321	0.026481508	alias hook3 - microtubule cytoskeleton organisation	cytoskeleton	Disregulated in t(1;11) translocation family (unpublished)	-	-
LOC100040996	1063807735	Q31LE/C57BL6E	-1.367679879	0.027184761	-	-	-	-	-
C9orf116RIK	1036605646	Q31LE/C57BL6E	-1.333825391	0.032505	-	-	-	-	-
B930007D08RIK	4230239	Q31LE/C57BL6E	-1.255795806	0.032907604	-	-	-	-	-
UBQK5	4060847	Q31LE/C57BL6E	-1.260628707	0.033503332	ubiquitin dependent protein catabolic process	nucleus	-	-	yes
E030019B13RIK	6180464	Q31LE/C57BL6E	-1.394462979	0.034275327	alias Gm6084 - function unknown	-	-	-	-
2310022M17RIK	6660037	Q31LE/C57BL6E	-1.506850106	0.035560759	-	-	-	-	-
SCL0002396_1_4	1033101211	Q31LE/C57BL6E	-1.262415401	0.036782498	alias POMC - neuropeptide signalling pathway	extracellular region	-	-	yes
A430073A17RIK	1001104111	Q31LE/C57BL6E	-1.428329342	0.036841374	alias ITGAV - cell differentiation, migration and adhesion	plasma membrane	-	-	yes
UBE2H	1022300368	Q31LE/C57BL6E	-1.276465032	0.038727842	post translational protein modification	-	-	-	yes
A830061P03RIK	1045680114	Q31LE/C57BL6E	-1.324981162	0.04077519	-	-	-	-	-
2410131K14RIK	1061101139	Q31LE/C57BL6E	-1.443368609	0.041549314	alias C12orf49	extracellular region	-	-	yes
AcBD6	4200446	Q31LE/C57BL6E	-1.25009473	0.044200047	AcetylCoA binding	-	-	-	yes
JUB	5860711	Q31LE/C57BL6E	-1.290644874	0.045109046	regulation cell migration, focal adhesion modification	lamellipodium, nucleus	-	-	yes
C130045G22RIK	1015005577	Q31LE/C57BL6E	-1.265949948	0.047166184	-	-	-	-	-
SCL0003103_1_727	1023407222	Q31LE/C57BL6E	-1.283771914	0.049660407	alias Rpa1 - single stranded DNA binding	DNA replication factor A complex	-	-	-

Table 4.9: New list of dysregulated genes post outlier removal. Continued next page…….

Gene_ID	Probe ID	Compaisson	Fold change	Pval	Function (and alias)	Localisation	Previous Implications in Major Mental Illness/ Disc1 Interaction	Overenrichment GO-Tree	IPA Top Canonical Pathways
TTC27	1090280	Q31L/C57BL6	-1.337021765	0.000206385	protein binding (assoc with PHLDA3)	-	-	-	-
IDE	670746	Q31L/C57BL6	1.286116186	0.000633312	proteolysis	cytoplasm	-	-	yes
2900017D05RIK	10269132	Q31L/C57BL6	1.26384156	0.000864185	-	-	-	-	-
TMEM181	10698028	Q31L/C57BL6	1.258161582	0.00090276	-	membrane	-	-	-
SCL0002547.1_9	10659067	Q31L/C57BL6	1.271361839	0.001952705	alias ENPP2 - alkylglycerophosphoethanolamine phosphodiesterase activity	plasma membrane	-	-	yes
CALML4	940154	Q31L/C57BL6	1.255336366	0.001976008	calcium ion binding	-	-	-	-
LOC100041504	5050338	Q31L/C57BL6	1.286625957	0.002199093	-	-	-	-	-
NDN	5670075	Q31L/C57BL6	1.369457738	0.002316554	axon extension and fasciculation	centrosome, nucleus	-	-	yes
CCL21B	4670176	Q31L/C57BL6	-1.284765843	0.003293117	immune reaction, negative regulation of myeloid cell differentiation	extracellular space	-	-	yes
DEFB11	2370440	Q31L/C57BL6	1.313907863	0.003804477	defensive response to bacteria	extracellular space	-	-	-
D230046H12RIK	10344074	Q31L/C57BL6	1.324622312	0.004092292	-	-	-	-	-
ERDR1	5890184	Q31L/C57BL6	-1.574492787	0.007874278	-	-	-	-	-
GLRX1	3520471	Q31L/C57BL6	-1.346717485	0.024463974	electron transport	cytoplasm	-	-	yes
5930418K15RIK	10485027	Q31L/C57BL6	1.269805525	0.035423175	Alias Tmem 181a	membrane	-	-	-
SYN1	4920301	Q31L/C57BL6	1.324684256	0.000427740	neurotransmitter secretion	synaptic vesicle membrane	-	-	-
GPR88	2450435	Q31L/C57BL6 Q31L/Q31LBupro	-1.1523167 1.26185129 -1.328703044	0.064165106 0.003597311 0.30492656	G-protein coupled receptor signalling pathway	membrane	-	-	-
1500015O10RIK	3130368	Q31L/C57BL6 Q31L/Q31LBupro	1.46036879 -1.559337696	0.031717451 0.045101114	-	extracellular space	-	-	-
CADPS	2320181	Q31L/Q31LBupro	1.472940643	0.000606467	positive regulation of calcium ion-dependent exocytosis	membrane/synapse	-	-	-
LRRC62	4120129	Q31L/Q31LBupro	1.295401397	0.000656892	alias Elfn2 - function unknown	cell surface	-	-	-
ATP6V0A1	1570082	Q31L/Q31LBupro	1.251864645	0.000819901	hydrogen ion transmembrane transporter activity	proton-transporting two-sector ATPase complex, proton-transporting domain	-	-	-
SLC17A7	7000037	Q31L/Q31LBupro	1.475263997	0.001110368	sodium-dependent phosphate transmembrane transporter activity	membrane	-	-	-
2310021P13RIK	4560500	Q31L/Q31LBupro	1.365219865	0.001230114	metal ion binding	-	-	-	-
SNPH	2680484	Q31L/Q31LBupro	1.300019077	0.001613664	negative regulation of endocytosis	synapse	-	-	-
SNCB	3360133	Q31L/Q31LBupro	1.261643853	0.001743792	negative regulation of neuron apoptosis	cytoplasm/synaptosome	-	-	-
CCDC131	1340603	Q31L/Q31LBupro	1.541851028	0.001838334	alias Zfc3h1 - function unknown	-	-	-	-
1110034O24RIK	10312003	Q31L/Q31LBupro	1.434397292	0.002295699	alias Tng02 - protein transporter activity	nucleus and cytoplasm	-	-	-
MDH1	6660358	Q31L/Q31LBupro	-1.252882315	0.002343656	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	cytoplasm/mitochondria	-	-	-
APPL2	2120280	Q31L/Q31LBupro	1.460833245	0.002911217	Rab GTPase binding	cytoplasm	-	-	-
ALG2	3990091	Q31L/Q31LBupro	1.36040233	0.003895943	transferase activity	membrane	-	-	-
PRKACB	4210170	Q31L/Q31LBupro	1.473405413	0.003923391	cAMP-dependent protein kinase activity	perinuclear region of cytoplasm	-	-	-
ATAD1	5550892	Q31L/Q31LBupro	-1.295074314	0.00420757	AtTP binding	mitochondria	-	-	-
SIRPA	130465	Q31L/Q31LBupro	1.303259846	0.00421875	protein phosphorylated amino acid binding	integral to plasma membrane	-	-	-
LY86	2230093	Q31L/Q31LBupro	-1.254045891	0.004298341	protein binding	extracellular	-	-	-
MAP3K12	2470373	Q31L/Q31LBupro	1.307423465	0.004424397	MAP kinase kinase kinase activity	growth cone	-	-	-
AKT1S1	1660048	Q31L/Q31LBupro	1.333706845	0.00512747	neuroprotection	cytoplasm	-	-	-
LOC100039346	7100594	Q31L/Q31LBupro	-1.270154886	0.005463788	-	-	-	-	-
A230046K03RIK	6130494	Q31L/Q31LBupro	1.32462991	0.005561975	endosome transport	-	-	-	-
CD300LD6RIK	10006010	Q31L/Q31LBupro	1.251581255	0.005743014	alias Ntrk2 - brain-derived neurotrophic factor binding	axon terminus/cytosol	-	-	-
KCTD13	2850064	Q31L/Q31LBupro	1.26400741	0.005966645	voltage-gated potassium channel activity	membrane	-	-	-
EG632248	10232072	Q31L/Q31LBupro	-1.397096367	0.006410098	alias Gm7083 - function unknown	-	-	-	-
AU040829	3830731	Q31L/Q31LBupro	1.253249689	0.006527212	Signal transduction	cytoplasm	-	-	-
EG433923	10237054	Q31L/Q31LBupro	-1.520328096	0.006826466	alias Gm6560 - function unknown	-	-	-	-
SRPK3	2480278	Q31L/Q31LBupro	1.312793599	0.007066275	protein serine/threonine kinase activity	-	-	-	-
ITM2A	460576	Q31L/Q31LBupro	-1.610267563	0.00732339	-	membrane	-	-	-
LOC100046330	10206060	Q31L/Q31LBupro	1.286564035	0.007425127	-	-	-	-	-
VDAC2	7000215	Q31L/Q31LBupro	-1.309506875	0.007665566	voltage-gated anion channel activity	mitochondrial membrane	-	-	-
HERC2	2970070	Q31L/Q31LBupro	1.269069318	0.00776275	acid-amino acid ligase activity	anaphase-promoting complex	-	-	-
WIPF3	10690057	Q31L/Q31LBupro	1.290483008	0.007976123	actin binding	cytoplasm	-	-	-
ANXA3	5570494	Q31L/Q31LBupro	-1.267040127	0.007994211	calcium-dependent phospholipid binding	neuron cell body	-	-	-
SERPINH1	6130014	Q31L/Q31LBupro	-1.369712226	0.008024345	serine-type endopeptidase inhibitor activity	endoplasmic reticulum	-	-	-
LOC383010	-	Q31L/Q31LBupro	-1.346132598	0.008355148	-	-	-	-	-
FCGR3	3840184	Q31L/Q31LBupro	-1.290381796	0.008370633	IgG receptor activity	membrane	-	-	-
RAB15	430070	Q31L/Q31LBupro	1.250677046	0.008566374	GTP binding	membrane	-	-	-
BZRAP1	5290487	Q31L/Q31LBupro	1.544367894	0.008707881	receptor activity	mitochondria/cytoplasm	-	-	-
LOC381608	10652034	Q31L/Q31LBupro	-1.484392602	0.008745689	-	-	-	-	-
TUBB4	5310681	Q31L/Q31LBupro	1.367138667	0.008761196	cell projection organization	intermode of axon	-	-	-
6430548M08RIK	2510019	Q31L/Q31LBupro	1.335090657	0.008800444	-	-	-	-	-
D6CR8	3450632	Q31L/Q31LBupro	1.277622958	0.008864602	double stranded RNA binding	nucleus	-	-	-
SLC25A3	6770112	Q31L/Q31LBupro	-1.669379953	0.008890718	symporter activity	mitochondrial inner membrane	-	-	-
LOC100047696	4280373	Q31L/Q31LBupro	-1.294089788	0.009288312	-	-	-	-	-

Table 4.9: New list of dysregulated genes post outlier removal. Continued next page…….

Gene_ID	Probe ID	Comparisson	Fold change	Pval	Function (and alias)	Localisation	Previous Implications in Major Mental Illness/ Disc1 Interaction	Overenrichment GO-Tree	IPA Top Canonical Pathways
ZNF512B	10590180	Q31U/Q31LBupro	1.334253789	0.009378357	-	-	-	-	-
BAT2	3840463	Q31U/Q31LBupro	1.345367434	0.009548975	-	cytoplasm/nucleus	-	-	-
RAB6	3840463	Q31U/Q31LBupro	-1.256410595	0.009622648	ATPase activator activity	membrane	-	-	-
2810455F06RIK	130136	Q31U/Q31LBupro	1.367705136	0.009866527	alias Ankrd16 - function unknown	-	-	-	-
1810034K20RIK	940465	Q31U/Q31LBupro	1.342868597	0.010499513	alias Mdp1 - protein tyrosine phosphotase activity	-	-	-	-
PSME3	2810537	Q31U/Q31LBupro	1.265592252	0.01069458	proteasome activator activity	nucleus	-	-	-
DCLK1	-	Q31U/Q31LBupro	1.263646639	0.010734637	protein serine/threonine kinase activity	-	-	-	-
HBXIP	2030279	Q31U/Q31LBupro	-1.320979932	0.010744574	-	cytoplasm	-	-	-
MED24	3170041	Q31U/Q31LBupro	1.268663864	0.010835803	transcription coactivator activity	nucleus	-	-	-
TMEM49	2900551	Q31U/Q31LBupro	-1.263416963	0.010845291	-	membrane	-	-	-
CELSR3	5270731	Q31U/Q31LBupro	1.262362401	0.010945999	G-protein coupled receptor activity	membrane	-	-	-
CD63	6660451	Q31U/Q31LBupro	-1.352460431	0.011019503	positive regulation of cell adhesion and endocytosis	membrane/lysosome	-	-	-
SFXN3	6350019	Q31U/Q31LBupro	1.253610221	0.011429574	cation transmembrane transporter activity	mitochondrial membrane	-	-	-
SLC25A14	4760180	Q31U/Q31LBupro	1.320029014	0.011576981	binding	mitochondrial inner membrane	-	-	-
EG432491	1062390273	Q31U/Q31LBupro	-1.62166016	0.011770772	alias Gm5426 - function unknown	-	-	-	-
EG433326	1004504403	Q31U/Q31LBupro	-1.510052575	0.01179615	alias Gm6529 - function unknown	-	-	-	-
LOC234582	1048101422	Q31U/Q31LBupro	-1.280620217	0.011987036	-	-	-	-	-
SNAPC4	2190184	Q31U/Q31LBupro	1.287678937	0.012153273	regulation of transcription	nucleus	-	-	-
LOC665386	1018501131	Q31U/Q31LBupro	-1.323337706	0.012439471	-	-	-	-	-
DCLK2	2510291	Q31U/Q31LBupro	1.282734494	0.012963873	protein serine/threonine kinase activity	cytoskeleton	-	-	-
TMEM63B	730397	Q31U/Q31LBupro	1.334956983	0.013002635	-	membrane	-	-	-
PIK3CB	3800600	Q31U/Q31LBupro	1.273740777	0.013297504	phosphatidylinositol-4,5-bisphosphate 3-kinase activity	phosphoinositide 3-kinase complex	-	-	-
AARS	1570279	Q31U/Q31LBupro	1.303701801	0.013489585	negative regulation of neuron apoptosis	cytoplasm	-	-	-
WSB1	870563	Q31U/Q31LBupro	1.260321196	0.013709072	intracellular signaling cascade	-	-	-	-
JOSD2	5080440	Q31U/Q31LBupro	1.27304134	0.013875407	-	-	-	-	-
TMED2	2370025	Q31U/Q31LBupro	-1.288453111	0.014531766	protein transport	cytoplasmic vesicle	-	-	-
SYP	2690168	Q31U/Q31LBupro	1.292429283	0.014930179	protein binding	presynaptic membrane	-	-	-
CD9	4730041	Q31U/Q31LBupro	-1.539618547	0.015107006	cell adhesion	apical plasma membrane	-	-	-
5730589K01RIK	4810286	Q31U/Q31LBupro	-1.318911212	0.015321082	alias tox4 - DNA binding	nucleus	-	-	-
SFRS6	60224	Q31U/Q31LBupro	1.289208983	0.015613741	-	-	-	-	-
FBXO25	460142	Q31U/Q31LBupro	-1.272848092	0.015803277	-	nucleus	-	-	-
TMEM25	2470575	Q31U/Q31LBupro	1.330444299	0.016030679	-	membrane	-	-	-
LOC100040961	-	Q31U/Q31LBupro	1.283113361	0.016048763	-	-	-	-	-
EIF4A2	1170494	Q31U/Q31LBupro	-1.596693951	0.017574461	ATP-dependent helicase activity	-	-	-	-
GRIPAP1	4590487	Q31U/Q31LBupro	1.282029134	0.017674854	-	endosome	-	-	-
SMG7	1031700 68	Q31U/Q31LBupro	1.276981206	0.017870803	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	cytoplasm	-	-	-
HPCAL4	2600484	Q31U/Q31LBupro	1.374978461	0.017923259	calcium channel regulator activity	-	-	-	-
KNDC1	4730452	Q31U/Q31LBupro	1.281659066	0.018754706	voltage-gated potassium channel activity	membrane	-	-	-
LOC639086	1061000 22	Q31U/Q31LBupro	1.346530621	0.019005152	-	-	-	-	-
NDUFS2	4860020	Q31U/Q31LBupro	-1.428753918	0.019335334	NADH dehydrogenase (ubiquinone) activity	mitochondrial inner membrane	-	-	-
C1QB	5910292	Q31U/Q31LBupro	-1.260365368	0.019748995	protein homodimerisation activity	complement component C1 complex	-	-	-
TMEM50A	2690440	Q31U/Q31LBupro	-1.754409244	0.019773065	-	membrane	-	-	-
CECR6	540672	Q31U/Q31LBupro	1.254799267	0.020047644	-	-	-	-	-
LOC100046853	1060206 00	Q31U/Q31LBupro	-1.345399142	0.020272221	-	-	-	-	-
CSF1R	2340110	Q31U/Q31LBupro	-1.269609666	0.020946871	transmembrane receptor protein tyrosine kinase activity	membrane	-	-	-
SELPLG	1770167	Q31U/Q31LBupro	-1.295746942	0.021083157	cell adhesion	integral to membrane	-	-	-
ARHGAP21	2100188	Q31U/Q31LBupro	1.357097817	0.021171068	Signal transduction	cytoskeleton/ golgi apparatus	-	-	-
CFP	2100619	Q31U/Q31LBupro	1.25821675	0.02124556	complement activation, alternative pathway	extracellular	-	-	-
RPL26	5220592	Q31U/Q31LBupro	-1.379684343	0.021642231	-	ribosome	-	-	-
LOC100043141	4850066	Q31U/Q31LBupro	-1.459691159	0.022013658	-	-	-	-	-
1200009C05RIK	3780136	Q31U/Q31LBupro	-1.259850785	0.022022118	alias Jkamp - function unknown	endoplasmic reticulum	-	-	-
BC005537	6700402	Q31U/Q31LBupro	-1.264007068	0.022116856	-	-	-	-	-
PFDN5	460020	Q31U/Q31LBupro	-1.327778941	0.0223772	protein foldong	-	-	-	-
AHDC1	5050747	Q31U/Q31LBupro	1.304088321	0.022663377	DNA binding	nucleus	-	-	-
FIBCD1	1500500	Q31U/Q31LBupro	1.259368945	0.022729884	signal transduction	integral to membrane	-	-	-
RNF130	4230731	Q31U/Q31LBupro	-1.346301764	0.023088816	metal ion binding	integral to membrane	-	-	-
FSCN1	3450463	Q31U/Q31LBupro	1.287086092	0.023574757	actin filiment binding	cytoskeleton	-	-	-
TSPAN3	7700102	Q31U/Q31LBupro	1.386411906	0.023759439	-	membrane	-	-	-
SMU1	5910128	Q31U/Q31LBupro	-1.286264666	0.023877116	-	cytoplasm/nucleus	-	-	-
BCAS1	2030711	Q31U/Q31LBupro	1.349060335	0.023926805	protein homodimerisation activity	cytoplasm	-	-	-
A230065H16RIK	1063505 00	Q31U/Q31LBupro	-1.512193196	0.024614406	-	-	-	-	-
HERPUD1	7100440	Q31U/Q31LBupro	-1.35716564	0.024615775	endoplasmic reticulum unfolded protein response	endoplasmic reticulum membrane	-	-	-
MOG	610364	Q31U/Q31LBupro	-1.424509264	0.024999823	cell adhesion	membrane	-	-	-

Table 4.9: New list of dysregulated genes post outlier removal. Continued next page…….

Gene_ID	Probe ID	Comparisson	Fold change	Pval	Function (and alias)	Localisation	Previous Implications in Major Mental Illness: Disc1 Interaction	Overenrichment GO-Tree	IPA Top Canonical Pathways
SV2B	3850717	031/U031LBupro	1.290436178	0.02505672	transporter activity	integral to synaptic vesicle membrane	-	-	-
GABRA2		031/U031LBupro	-1.270411515	0.025975777	extracellular ligand-gated ion channel activity	membrane	-	-	-
LOC100046088	3170309	031/U031LBupro	1.297525146	0.026084132	-	-	-	-	-
PRKCB	8700019	031/U031LBupro	-1.454775692	0.026093978	calcium channel regulator activity	cytoplasm/nucleus	-	-	-
LOC100043919		031/U031LBupro	1.25480125	0.026331476	-	-	-	-	-
LOC668706	7100750	031/U031LBupro	-1.319251063	0.026370728	-	-	-	-	-
LOC228017	1015701 42	031/U031LBupro	-1.290116649	0.026640061	-	-	-	-	-
SPCS1	730373	031/U031LBupro	-1.831934921	0.02681139	peptidase activity	integral to endoplasmic reticulum membrane	-	-	-
LOC270186	100940739	031/U031LBupro	-1.362543385	0.026825898	-	-	-	-	-
2310007G05RIK	1038403332	031/U031LBupro	1.266373486	0.026848584	alias Foxp4 - regulation of transcription (negative)	nucleus	-	-	-
D11BWG0434E	1740450	031/U031LBupro	-1.367229165	0.026874396	alias Rnask - endonuclease activity	membrane	-	-	-
PITPNM1	4540528	031/U031LBupro	1.266128196	0.026786966	phospholipid binding	Golgi apparatus	-	-	-
LOC100047226	1017400 10	031/U031LBupro	-1.272469809	0.026892384	-	-	-	-	-
AI480566	540315	031/U031LBupro	1.401836169	0.026976758	alias Scaf1 - protein domain specific binding	nucleus	-	-	-
E130301F21RIK	1011907 07	031/U031LBupro	-1.262178742	0.029109133	-	-	-	-	-
A230057G18RIK		031/U031LBupro	1.515395165	0.029249697	alias Mist - retinal rod development	-	-	-	-
SFRS16		031/U031LBupro	1.27837988	0.029609201	mRNA processing	nucleus	-	-	-
PALM	4120068	031/U031LBupro	1.26692154	0.029881953	D3 dopamine receptor binding	membrane	-	-	-
PDXP	5050204	031/U031LBupro	1.253300118	0.03036076	pyridoxal phosphatase activity	-	-	-	-
ZNHIT3	1037904 11	031/U031LBupro	-1.253935773	0.031225107	-	-	-	-	-
RPS3	1570017	031/U031LBupro	-1.486896236	0.031715619	structural constituent of ribosome	ribonucleoprotein complex	-	-	-
B0021381	5550180	031/U031LBupro	1.32739298	0.031760791	alias Fam193b - function unknown	-	-	-	-
SPNB3	6660088	031/U031LBupro	1.332639009	0.031848026	actin binding	cytoplasm	-	-	-
RBBP7	430113	031/U031LBupro	-1.269086805	0.033373537	transcription repressor activity	NuRD complex	-	-	-
C1QC	5700131	031/U031LBupro	-1.275374376	0.0334906	innate immune response	extracellular	-	-	-
EIF5A	1500069	031/U031LBupro	-1.313100674	0.03352255	translation elongation factor activity	cytoplasm	-	-	-
ARPC1A	70672	031/U031LBupro	-1.405655195	0.033815563	actin binding	cytoskeleton	-	-	-
MAP2K4	5130133	031/U031LBupro	1.269027081	0.033952024	mitogen-activated protein kinase kinase kinase binding	cytoplasm/nucleus	-	-	-
LY6C1	1990446	031/U031LBupro	-1.447804537	0.0346498	-	external side of plasma membrane	-	-	-
E0620119		031/U031LBupro	-1.588677	0.034812186	alias Gm6128 - function nuknown	-	-	-	-
PCOLCE2	5380170	031/U031LBupro	-1.277510815	0.035081126	heparin binding	extracellular region	-	-	-
SEPW1	3800079	031/U031LBupro	1.520190192	0.035219806	oxidoreductase activity/selenium binding	plasma membrane	-	-	-
DYSWSU169E	2350022	031/U031LBupro	1.307060321	0.036151216	signal transduction	cytoskeleton	-	-	-
XBP1	3840594	031/U031LBupro	-1.348689676	0.036473778	transcription factor activity	nucleus	-	-	-
MT3	1450537	031/U031LBupro	1.736859587	0.036847965	negative regulation of neurogenesis	synaptic vesicle	-	-	-
CSMD2	1006102 79	031/U031LBupro	1.261220199	0.037090205	-	-	-	-	-
LGMN	3610301	031/U031LBupro	-1.410090424	0.037639765	cysteine-type endopeptidase activity	late endosome	-	-	-
MYO9B		031/U031LBupro	1.254007817	0.037968304	actin cytoskeleton organization	actin filament	-	-	-
SLC6A8	3940341	031/U031LBupro	1.330611352	0.038456583	choline transmembrane transporter activity	membrane	-	-	-
SIR7	5890632	031/U031LBupro	1.287653122	0.03897002	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in linear amides	nucleus	-	-	-
SUCLG1	1850062	031/U031LBupro	-1.296443326	0.038993732	ATP citrate synthase activity	mitochondrial inner membrane	-	-	-
ZBTB38	670056	031/U031LBupro	1.257901811	0.039006587	metal ion binding	-	-	-	-
LOC100044163	3140131	031/U031LBupro	1.276341161	0.039330297	-	-	-	-	-
RHOBTB3	4010667	031/U031LBupro	1.41954601	0.039890603	ATP/GTP binding	Golgi apparatus	-	-	-
1810015C11RIK	1029076 47	031/U031LBupro	1.315061265	0.040159672	-	-	-	-	-
HBB-B1	4050717	031/U031LBupro	-2.444043302	0.043076986	oxygen transporter activity	haemoglobin complex	-	-	-
SLC2A3	1990377	031/U031LBupro	1.26621232	0.043196054	D-glucose transmembrane transporter activity	membrane	-	-	-
TUBA1C	5200441	031/U031LBupro	-1.280611985	0.043496008	structural molecule activity/GTP binding	microtubule	-	-	-
HP1BP3	2650575	031/U031LBupro	1.347275102	0.0442537	DNA binding	nucleosome	-	-	-
TRF	1050193	031/U031LBupro	-1.693098655	0.04490783	ferric iron transmembrane transporter activity	endocytic vesicle	-	-	-
E0630409	4210184	031/U031LBupro	-1.285728216	0.045244247	alias Gm7035 - function unknown	-	-	-	-
CAB39L	5670040	031/U031LBupro	-1.258878145	0.045975174	binding	-	-	-	-
ARF5		031/U031LBupro	-1.447863465	0.046999961	small GTPase mediated signal transduction	golgi apparatus	-	-	-
LOC100047252	1570102	031/U031LBupro	-1.259785945	0.04737966	-	-	-	-	-
AHSA1	2030520	031/U031LBupro	-1.284498923	0.047643489	chaperone binding	cytoplasm	-	-	-
KCTD6	4070435	031/U031LBupro	-1.354990363	0.048004508	voltage-gated potassium channel activity	membrane	-	-	-
LOC100046623	1056707 06	031/U031LBupro	-1.257299439	0.048444338	-	-	-	-	-
LOC545208	1056604 64	031/U031LBupro	-1.268209721	0.048728115	-	-	-	-	-
ZIC1	6700113	031/U031LBupro	-1.670454481	0.048900455	DNA/protein binding	nucleus	-	-	-
VPS29	6808087	031/U031LBupro	-1.364379322	0.049062716	phosphoserine phosphatase activity	membrane/cytoplasm	-	-	-
LOC268689	1038400 92	031/U031LBupro	-1.740786126	0.049087607	-	-	-	-	-
DRG1	1660372	031/U031LBupro L100P/A100P/Roli	-1.28906357	0.04891235	GTP binding	intracellular	-	-	-
			1.522828927	0.017657664					

Table 4.9: New list of dysregulated genes post outlier removal. Column 1 and 2 show the gene symbol (Gene_ID) and probe ID, column 3 shows the comparison in which it is dysregulated, columns 4 and 5 show fold-change and p-value, column 6 gives function GOterms and possible alias', column 7 gives localisation where possible, column 8 further evidence to support validation from previously published studies, column 9 and 10 state whether the gene was overexpressed in GOTree analysis and/or IPA analysis.

Some probes were found to be differentially expressed in more than one comparison (figure 4.8)

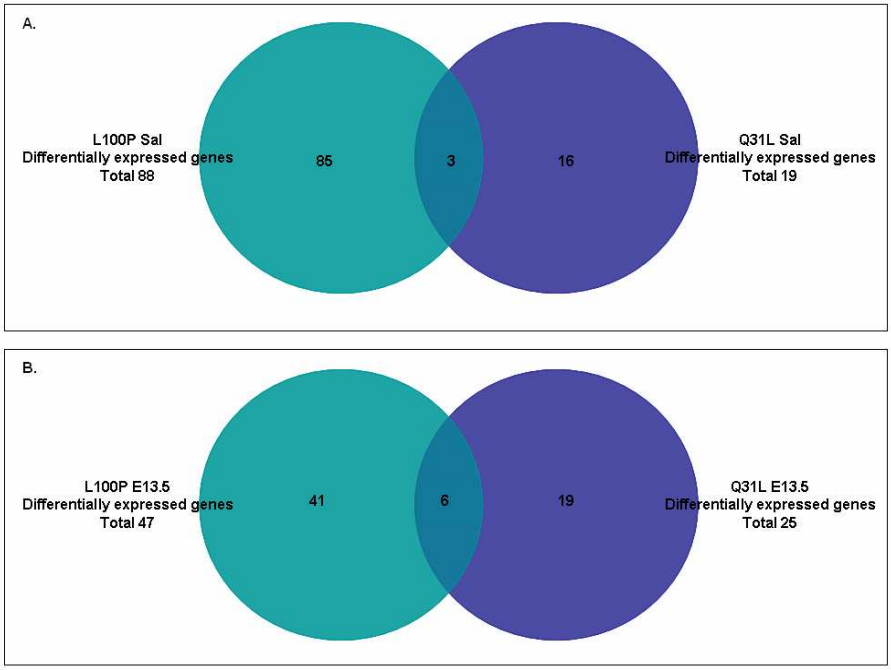

Figure 4.8: Venn diagram of significant probes in the genotype comparisons. Pane A shows significant probes in both adult comparisons while pane B shows significant probes in both embryonic comparisons. The total number of differentially expressed genes are shown. Within the circles of the Venn diagram are the number of genes unique to each comparison and, where the circles overlap, the number common to the comparisons.

In addition, there were eight genes that showed correction with drug treatment, and a further four genes that had borderline fold-change and p-value for correction with drug treatment. Bupropion treatment of the Q31L saline mouse corrected three genes back to the expression level of the wild type, clozapine treatment of the L100P mouse corrected one gene back to the expression level of the wild type, and rolipram treatment of the L100P mouse corrected five genes, and was close to correction of four genes back to a level not significantly different to that of the C57BL/6J mouse.

4.4.2 New Gene Ontology, Pathway analysis and Comparisons with previous study

Of the 88 genes shown to be differentially expressed in the L100PSaline group, 62 were upregulated and 26 were downregulated. GOTree analysis highlighted 11 biological processes over-enriched in the sample-set (table 4.10)

Biological process	Observed	Expected	P value	Genes
cell-cell signaling	5	0.81	0.0012032	Egr2, Egr3, Gria2, Nrxn, Syn2
transmission of nerve impulse	5	0.55	0.0002151	Egr2, Egr3, Gria2, Nrxn, Syn2
synaptic transmission	4	0.46	0.0010831	Egr3, Gria2, Nrxn, Syn2
neurotransmitter secretion	2	0.11	0.0058543	Nrxn, Syn2
secretory pathway	7	0.57	0.0000013	Cplx2, Folr1, Napb, Nrxn, Exoc4, Syn2, Sec24d
vesicle docking during exocytosis	2	0.05	0.001013	Cplx2, Exoc4
regulated secretory pathway	3	0.13	0.000284	Cplx2, Nrxn, Syn2
vesicle-mediated transport	5	1.03	0.0034956	Cplx2, Gria1, Napb, Exoc4, Sec24d
vesicle docking	2	0.05	0.0011376	Cplx2, Exoc4
peripheral nervous system development	2	0.05	0.0011376	Egr2, Egr3
secretion	7	0.75	0.0000078	Cplx2, Folr1, Napb, Nrxn, Exoc4, Syn2, Sec24d

Table 4.10: Table of biological processes over-enriched in the L100PSaline group. Column 1 shows the biological process, column 2 the number of genes in the dataset involved in the process, column 3 the number of genes expected in a dataset, column 4 the p-value of the finding, and column 5 the genes found.

None of the other groups showed overenrichment of GO categories. IPA analysis revealed a top network score of 48 “cellular development” for the L100P adult comparison, with 21 dysregulated genes appearing in the network (figure 4.9). The top canonical pathway for L100P was synaptic long-term depression (p= 0.005). There was now a significant canonical pathway for the Q31L analysis (p=0.01, figure 4.10 panel B), but no significant network of genes was identified. The L100P embryo pathway analysis changed very little, but the Q31L embryonic network analysis was greatly changed. For the L100P embryo the top network identified, “gene expression”, gave a score of 25, and for the Q31L embryo the top network “cell signalling” gave a score of 35. A score of 25 represents a 1.0E-24 probability of the genes within the network grouping by chance. While for the L100P embryo notch signalling remained the top canonical pathway (p=0.006), for the Q31L embryo this was replaced by caveolar-mediated endocystosis (p=0.001), pushing notch signalling into second spot (figure 4.10).

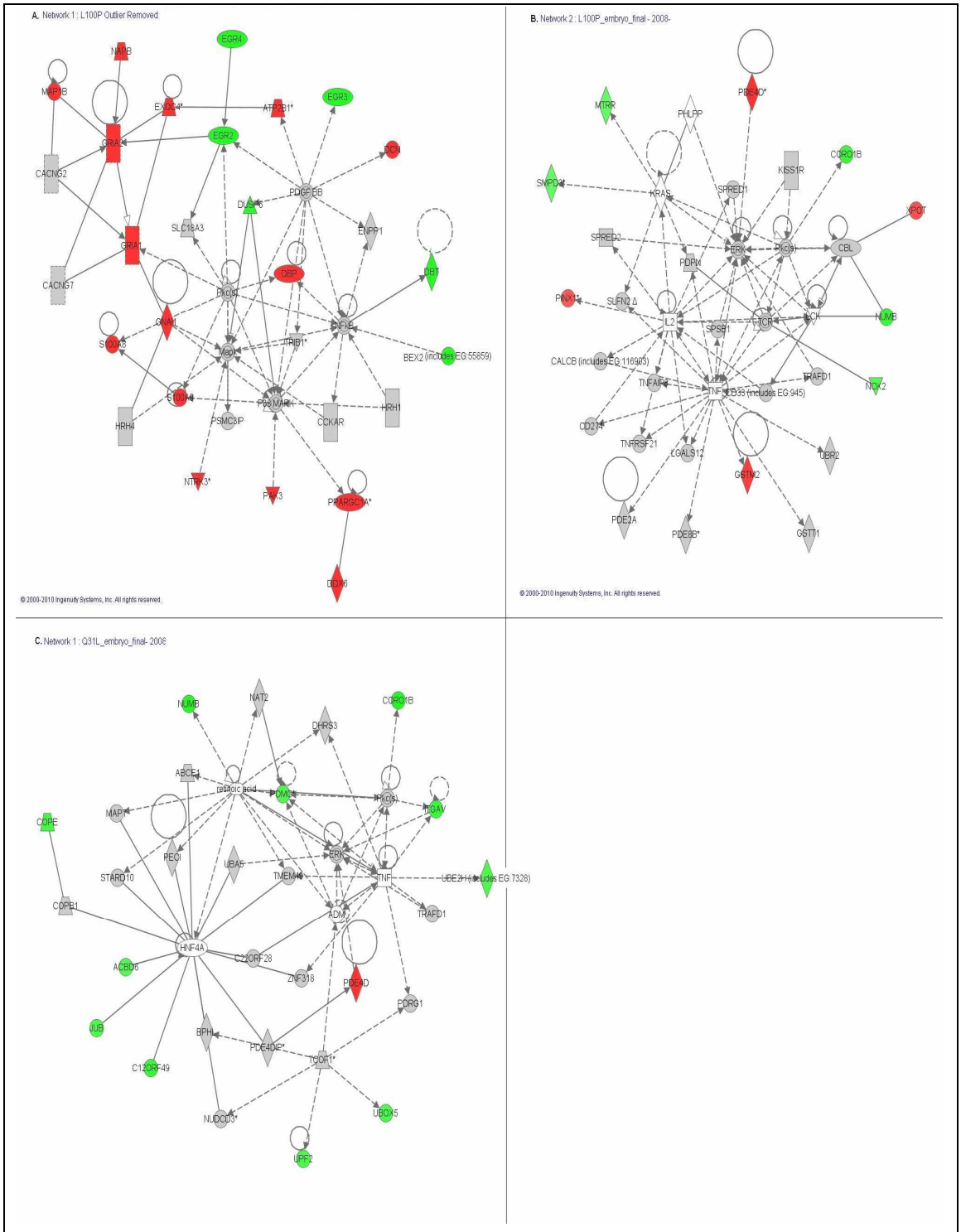

Figure 4.9: Ingenuity Pathway Analysis. Panel A shows top L100P adult network "cellular development". Panel B shows the top L100P embryo network "gene expression" and Panel C the top Q31L embryo network "cell cignalling". Those in green are downregulated, and those in red upregulated. Genes shaded grey do not appear in the dysregulated gene, but are expressed above background in the dataset.

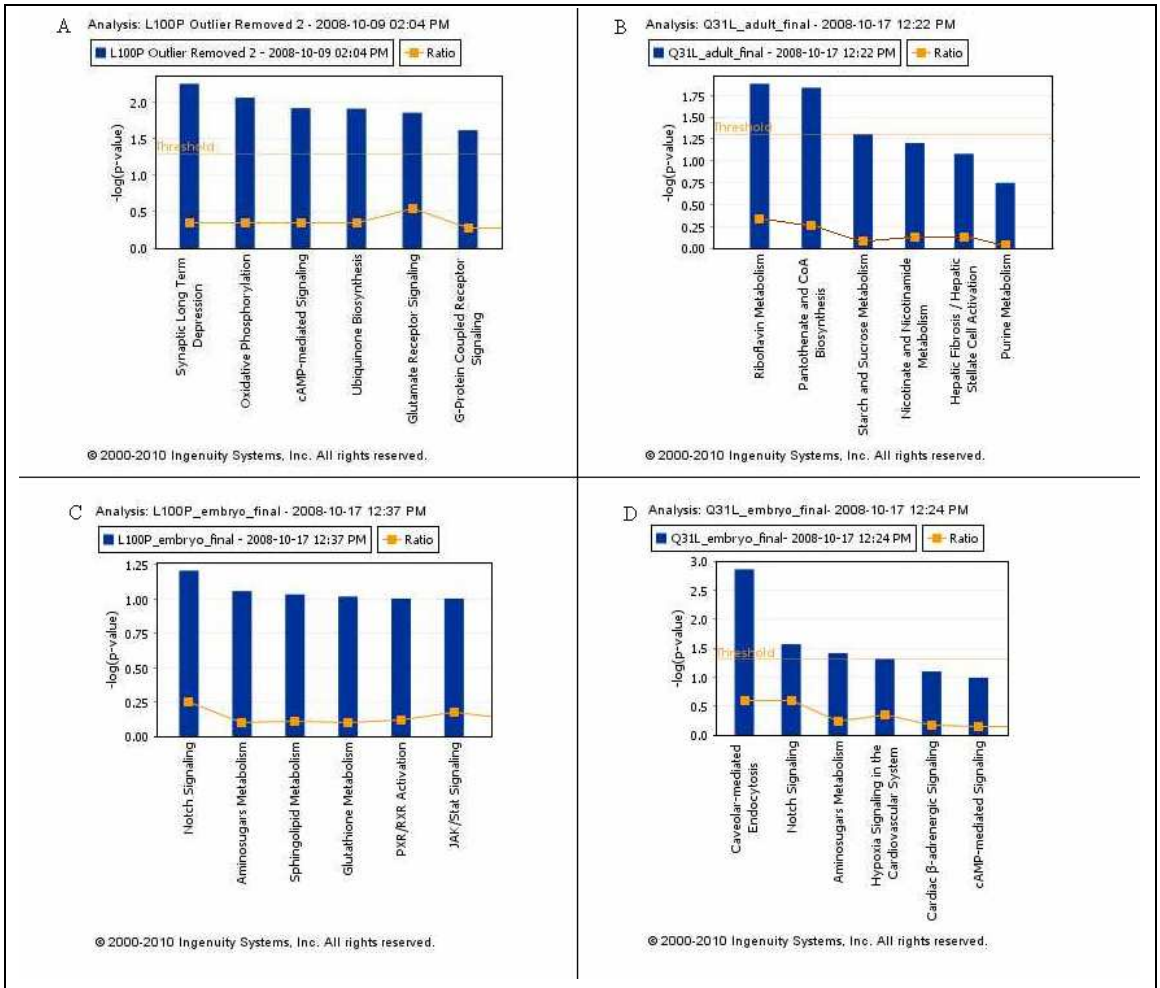

Figure 4.10: Canonical pathways for IPA analysis on new dysregulated gene list. Panel A shows the top six pathways in the L100P adult group, panel B the top six pathways in the Q31L adult group, panel C the top six pathways in the L100P embryonic group, and panel D the top six pathways in the Q31L embryonic group. *Threshold* bar shows cut-off point of significance $P < 0.05$. The yellow 'ratio' line shows the *ratio* of genes in network to total number of genes in *canonical* pathway.

There were also eight genes found to overlap with previous studies of major mental illness, with one gene being present in two of the studies included in the analysis (table 4.11). Probability distributions consider this a significant value ($p<0.001$). Studies used for comparison are described in section 4.2.3.

Gene	Study	Gain/Loss	Function
DUSP6	t1:11 translocation carriers (Millar et al unpublished)	Down-regulated, Up-regulated	MAP kinase tyrosine/serine/threonine phosphatase activity
F5	t1:11 translocation carriers (Millar et al unpublished)	Up-regulated, Down-regulated	blood coagulation/circulation
HOOK3	t1:11 translocation carriers (Millar et al unpublished)	Down-regulated, down-regulated	microtubule cytoskeleton organisation
NRXN1	t1:11 translocation carriers (Millar et al unpublished) Kirov et al (2009)	Up-regulated, Up-regulated, CNV duplication	synaptic transmission
NRXN3	t1:11 translocation carriers (Millar et al unpublished)	Up-regulated, Up-regulated	synaptic transmission
MLL5	Walsh et al (2008) Schizophrenia CNVs (cases)	Up-regulated, CNV duplication	metal ion binding, regulation of transcription
PDE4D	Camargo et al (2006) DISC1 interactome	Up-regulated -	cAMP catabolism, signal transduction
FBXO41	Camargo et al (2006) DISC1 interactome	Up-regulated -	ubiquitin dependent protein catabolic process

Table 4.11: Overlap with previous studies. This table shows the overlaps between genes from the current microarray study and large genome wide studies previously published from individuals with schizophrenia and other major mental illness. All genes with overlap were found within the L100P non-drug treated (saline) group. Column 1 shows the gene name, column 2 the comparative previously published evidence, column 3 the direction of change (with this study first and comparative study second) and column 4 the gene function/location.

4.5 New list of genes for qRT PCR

Genes were selected for validation by qRT PCR based on fold-change, p-value, GOTree analysis, previous implications in mental illness research and overlaps with current studies of the human (t1:11) translocation carriers (Table 4.12). Initially, genes were sorted based on p-value and then fold change, and scored by where they fell within the list (ie those in the top 10% based on p-value were given a score of 10, top 11-20% scored 9 etc). Genes were then scored on whether they appeared in the 'over-enriched' functions based on GOTree analysis and/or Ingenuity pathway analysis. Finally, genes were scored based on whether they had previous implications in major mental illness or overlapped with the human translocation carrier data. Scores were totalled and the top 10% of genes from each comparison (now based on the total combined score from p-value, fold change, GOTree analysis, IPA and overlaps with other studies) were carried forward for further validation.

Gene	Comparisson	Array FC	Array Pval	GO Terms	Evidence
1500015O10RIK	Q31L/C57BL6 Q31L/Q31LBupro	1.46 -1.56	0.03 0.045	unknown	-
Arc	L100P/C57BL6	-1.35	0.01	actin binding post synaptic membrane	Over-enriched Gotree category
Atp5b	L100P/C57BL6 L100P/L100PRoli	1.23 -1.37	0.02 0.04	proton-transporting ATP synthase complex, coupling factor F(o)	Over-enriched Gotree category
Bex2	L100P/C57BL6 L100P/L100PRoli	.-1.3 1.22	0.00004 0.06	apoptosis/cell cycle	IPA pathway
Cdh11	L100PE/C57BL6E	1.55	0.049	cell adhesion	IPA pathway
CORO1B	L100PE/C57BL6E Q31LE/C57BL6E	.-1.61 -1.72	0.026 0.012	actin binding	-
Cplx	L100P/C57BL6	1.41	0.01	neurotransmitter transport	Over-enriched Gotree category
Cstf1	L100PE/C57BL6E	1.64	0.047	mRNA processing	IPA pathway
DUSP6	L100P/C57BL6	-1.28	0.018	MAP kinase tyrosine/serine/threonine phosphatase activity	Over-enriched Gotree category, IPA pathway, Disregulated in t1:11 translocation family (unpublished)
EGR2	L100P/C57BL6	-1.26	0.001	brain segmentation, motor axon guidance, schwann cell differentiation	Over-enriched Gotree catagory, IPA pathway, Yamada et al 2006
EGR3	L100P/C57BL6	-1.26	1.09E-05	neuromuscular synaptic transmission	Over-enriched Gotree catagory, IPA pathway, Yamada et al 2007
EGR4	L100P/C57BL6	-1.36	0.005	regulation of transcription, DNA-dependent	Over-enriched Gotree catagory, IPA pathway, Yamada et al 2008
Exoc4	L100P/C57BL6	1.26	9.25E-05	vesicle docking during exocytosis	IPA pathway
FBXO41	L100P/C57BL6	1.33	0.003	ubiquitin dependent protein catabolic process	Camargo et al 2006
GPR88	Q31L/C57BL6 Q31L/Q31LBupro	1.26 -1.32	0.003 0.03	G-protein coupled receptor signalling pathway	-
GRIA1	L100P/C57BL6	1.3	0.0004	receptor internalisation, ion transport	Over-enriched Gotree catagory, IPA pathway
GRIA2	L100P/C57BL6	1.27	0.01	synaptic transmission	Over-enriched Gotree catagory, IPA pathway
Mast3	L100P/C57BL6 L100P/L100PRoli	.-1.19 1.26	0.001 0.03	protein amino acid dephosphorylation	-
MLL5	L100P/C57BL6	1.29	0.002	chromatin modification, cell cycle arrest	Over-enriched Gotree category, Walsh et al 2008
NAPB	L100P/C57BL6	1.29	0.001	syntaxin binding	Over-enriched Gotree catagory, IPA pathway
Ndfip	L100P/C57BL6 L100P/L100PRoli	1.45 -1.72	0.009 0.019	protein binding	-
NDN	Q31L/C57BL6	1.36	0.02	axon extrension and fasciculation	IPA pathway
NRXN1	L100P/C57BL6	1.35	0.001	synaptic transmission	Over-enrichment Gotree pathway, Kirov et al 2008; Need et al 2009; Vrijenhoek et al 2008; Rujescu et al 2009
NRXN3	L100P/C57BL6	1.25	0.022	calcium channel regulator activity	Novak et al 2009; disregulated in t1:11 translocation (unpublished)
NUMB	L100PE/C57BL6E Q31LE/C57BL6E	.-1.65 -1.71	0.039 0.023	axonogenesis, forebrain development	-
PAK3	L100P/C57BL6 L100P/L100PRoli	1.35 -1.28	0.0002 0.049	kinase activity	Over-enriched Gotree catagory, IPA pathway, Bienvenu et al 2000; Donelly et al 1996; Peippo et al 2007
PDE1A	L100P/C57BL6	1.27	0.002	calcium- and calmodulin-regulated 3',5'-cyclic-GMP phosphodiesterase activity	Over-enriched Gotree category
PDE4D	L100PE/C57BL6E Q31LE/C57BL6E	1.94 1.42	0.009 0.007	cAMP catabolism, signal transduction	Camargo et al 2006; Millar et al 2005
POP4	L100P/C57BL6 L100P/L100PRoli L100P/L100PCloz	.-1.62 1.7 1.73	0.005 0.05 0.05	rRNA processing/cleavage	-
PPARGC1	L100P/C57BL6	1.28	0.001	positive regulation of transcription from RNA polymerase II promoter	Over-enriched Gotree catagory, IPA pathway
PRM3	L100PE/C57BL6E Q31LE/C57BL6E	.-1.53 -1.31	0.011 0.034	cell differentiation, mitotic chromosome condensation	-
Prmt6	L100PE/C57BL6E	-2.24	0.007	methyltransferase activity	IPA pathway
RANBP9	L100P/C57BL6 L100P/L100PRoli	1.14 -1.33	0.07 0.04	RAN GTPase binding	-
SYN1	Q31L/C57BL6 Q31L/Q31LBupro	1.32 -1.15	0.0004 0.064	neurotransmitter secretion	-
Syn2	L100P/C57BL6	1.3	0.002	neurotransmitter secretion	Over-enriched Gotree category
WDFY1	Q31L/C57BL6 L100P/C57BL6	1.5 1.46	4.9E-10 2.85E-10	metal ion binding	Over-enriched Gotree category

Table 4.12: List of genes to be analysed by QT RT PCR. Column1 gives the gene name, column 2 the comparison in which it is dysregulated, column 3 the fold change from the array, column 4 the p-value from the array, column 5 the GeneOntology terms for that gene, and column 6 extra evidence for further analysis.

In total 36 genes were selected for further analysis, 20 from the non-drug treated adult mice, ten that showed correction, or near to correction, with drug treatment, and six from the embryonic stages. qRT PCR analysis of these genes will be described in the following chapter.

4.6 Conclusions

This chapter describes the analysis and significance of the results obtained from a genome-wide expression study as described in chapter 3. Genes that were found to be differentially expressed were weighted on their functional significance (based on gene ontology and pathway analysis) and previous implications in major mental illness research. Genes which showed correction by drug treatment were also considered.

The initial analysis revealed a total of 835 genes that were differentially expressed across the sample set. As the one outlier pool gave erratic and unpredictable expression values, being neither consistently higher nor lower than the other samples, this likely explains why the initial analysis did not identify this pool as an outlier, as the effect was muted by combining all genes together. Thus when the average expression for each pool was compared no pool was initially identified as an outlier. This shows the need to take comparable biological groups and analyse the raw data of a subset of genes to ensure outliers are not missed in future.

Secondary analysis revealed a total of 368 genes differentially expressed across the sample set. In hindsight, secondary analysis could have been avoided by checking a random subset of genes from the outset, rather than after the analysis had been carried out.

The comparison between the L100P adult and the C57BL/6J adult mice identified 88 genes that were differentially expressed, 62 of which were over-expressed and 24 under-expressed in the L100P adult samples. The Q31L bupropion treated mouse showed the

highest levels of differential expression in the drug treated groups with 91 genes over-expressed and 95 genes under-expressed compared to non-drug treated adults. Only three of these genes were also found to be dysregulated in the Q31L non-drug treated mouse and it is likely the high numbers are a factor of the drug treatment and not the Q31L genotype. The L100P rolipram treated mouse gave the most genes which showed correction under pharmacological influence. Of the 32 genes identified, seven showed correction of the L100Pnon-drug treated disregulation. Correction of expression was defined as returning that gene to a level not significantly different (FC+/- 1.3, $p<0.05$) to that of C57BL/6J controls.

In the embryonic groups, L100P identified 45 differentially expressed genes, while Q31L identified 25. Four genes were identified in both groups, PRM3, Coro1B, Numb and PDE4D.

Selection of genes for further analysis was performed in multiple ways. Genes were first ranked in order of fold-change and p-value then GOTree analysis was performed to determine which biological processes, cellular components and molecular functions were or were not overrepresented in the differentially expressed gene lists generated from the microarray analysis. This not only facilitates direct comparison with other studies of major mental illness due to the simple and uniform characterizations, but also helped to "compartmentalize" the data and identify promising genes for further analysis. The GOTree analysis revealed over-enrichment of genes involved in cell-cell signaling, transmission of nerve impulses and neurotransmitter secretion in the L100P non-drug treated (saline) adult dataset. Among these genes are the two *Neurexins* (*Nrxn1* and *Nrxn3*), *Gria1* and *Gria2, Erg2, Egr3* and *Egr4*. Both *neurexin* genes and *Gria1* and *Gria2* were found to be over-expressed in the L100P adult samples, while the three *Egr* genes were found to be under-expressed in the sample set.

Ingenuity pathway analysis was also used to identify dysregulated genes that grouped together to form established networks and pathways. Nine genes showed both overenrichment in GO categories and were present in the top statistically significant IPA

network for the relevant group. One of the potential pitfalls of Ingenuity analysis for this study is the relevance of known networks to major mental illness. Much of the data currently held refers to networks established by studies of cancer genetics and diseases with simple genetic components. While these networks currently provide good points of reference and identify potential genes for further analysis, many connections are indirect. With continuing genetic studies being added to the existing datasets this is constantly improving but for this study IPA was used purely for identifying potential genes for further analysis and not to make direct claims to gene networks involved in major mental illness at this stage. Because the Ingenuity database is constantly being updated with new datasets, it is a dynamic tool and justifies repeat analysis in the future to determine whether new networks are identified with the same experimental dataset.

Comparing the dataset from this study to selected previously published studies allowed me to not only identify genes for further analysis, but to determine how this study sat in relation to previous work, and to gauge the relevance of these two ENU mouse models to major mental illness research. Recent work has investigated the role of copy number variants (CNVs) in neurodevelopmental disorders (reviewed in Kirov 2010) [82]. As the aetiology of schizophrenia and other major mental illness is so varied, with multiple genetic components identified, the general consensus is that multiple small components are likely to be responsible, and as CNVs account for a substantial proportion of human genetic variation, are likely to play an important role. With this in mind I compared my dataset with three recently published studies of CNVs in major mental illness to determine whether they were comparable. Only two genes were present in both the dysregulated gene list from this study and copy number variants in the previous studies. *Mll5*, a transcription regulator, was present as a duplication in schizophrenia cases from Walsh 2008 [83], and was upregulated in the L100P non-drug treated adult group. MLL5 has not been reported to interact directly with DISC1 [132], but DISC1 is well established as interacting with other kinases and modulating signaling pathways. For example, DISC1 interacts with PDE4B to dynamically regulate cAMP signaling, and also interacts with the cAMP response element-dependent transcription factor, ATF4, possibly binding to chromatin remodelling factors, such as SMARCE1 [132]. PDE4s are

orthologous to the Drosophila *Dunce*, which is involved in learning and memory and known to affect synaptic plasticity, which in turn requires alteration of gene expression profiles [170]. These interactions with transcription factors, together with the fact that DISC1 localises to the nucleus [172, 173] is consistent with a role for DISC1 in transcriptional regulation. *Nrxn1* was also found to be up-regulated in the L100P non-drug treated adult group, with a CNV deletion present in a schizophrenia case and affected sibling [217]. The deletion spans the promoter and first exon of the gene and partially overlaps with previous deletions identified in autism [76] and mental retardation [239].

I also had access to data from recently completed microarray studies on the DISC1 translocation (t1:11) family that allowed an almost direct comparison to be made between the mouse and human model. Five genes (*Dusp6, F5, Hook3, Nrxn1* and *Nrxn3*) were present in both studies, with *Hook3, Nrxn1* and *Nrxn3* dysregulated in the same direction in both arrays. These overlapping results between the mouse and human sample subjects is promising as it provides further evidence of the validity of the animal model to this system. One major flaw of this comparison is the two arrays used different tissues as the primary RNA source. While it was possible to use brain tissues from the ENU mutant mouse, the human samples were from a blood-derived cell line. Blood-derived (lymphoblastoid) cell lines do not express all genes and only a sub-set of those expressed in the brain[240]. This could obviously result in some genes not being present in the human dataset by virtue of the original tissues used, and not because they are not expressed, or potentially dysregulated in the human condition.

This study had obvious limitations and weaknesses, not least of all the presence of an outlier group in the initial analysis. As described previously, the Illumina built in controls did not identify this one pool as an outlier, highlighting the need to assess a subset of genes raw data across a group for complete confidence. A number of the genes identified as differentially expressed on the array were from a cDNA bank, with little or no functional information available. This made it difficult to determine the relevance of these predicted sequences to major mental illness and as such most were not chosen for

further analysis. It is possible that the lack of information available for many genes biased the selection of the subset for further analysis. While the methods used for microarray analysis described in this section have been widely used, it is important to note that other methods of data analysis are equally popular and different statistic procedures will give different lists of genes. The outcome of the analysis is obviously dependent on the starting data-set, thus testing for differential expression and then filtering on fold-change (rather than filtering by fold-change then potential differential expression as described here) would potentially lead to a different varience distribution (as more probes would be used to estimate the varience) and a different test statistic. This would potentially generate a different gene list from the same data. This has previously been addressed [241] with the conclusion that care should be taken when interpreting any results that rely on the structure of the starting dataset. It may be suggested that the dataset be reanalyzed in the future using the various methods available and only the genes which are consistant through all analyses be retained for further follow-up work. Non-parametric tests, Baysian models and t-tests (as used in this study) are all used to generate lists of differentially expressed genes from normalized microarray datasets [242]. Furthermore, once a list of genes is generated, there are multiple methods of describing changes in expression and identifying functional significance. Heirarchical clustering methods group genes by expression levels, but are of little use in determining functional significance [243]. Pathway analysis (as used here) is useful as it groups genes by GOTree classifications and so can be useful in identifying common functions that may be affected by gene expression changes [243]. Neither of these methods, however, address the issue of false-positives and erroneous artifacts, which may lead to genes being studied further with no viable results. Bertsch *et al* (2005) propose to narrow the list of genes using Convergent Functional Genomics (CFG), which uses Baysian methods to cross-validate datasets between arrays and linkage data, with the aim of reducing uncertainty and removing false positives [244].

The multiple stage selection process was employed to select the potentially most interesting and viable genes for further analysis, based on current knowledge of gene ontology, gene pathways and previous genetic work on schizophrenia and major

depression. However, there are multiple genes identified in the initial analysis which will not be investigated further at this stage. Ideally all genes identified would be validated using qRT-PCR but constraints on finances and time meant it was necessary to prioritise the most interesting and potentially viable genes for further study.

In summary, this study has identified several putative candidate genes in both embryonic and adult stages. Correction of gene expression through drug administration has also been shown and is worthy of further investigation.

Chapter 5
Quantitative real-time PCR analysis of candidate differentially expressed genes in the L100P and Q31L mutant mice

5. Quantitative real-time PCR analysis of candidate differentially expressed genes in the L100P and Q31L mutant mice

5.1 Introduction

For the past decade researchers have been using qRT-PCR as a robust and reliable measure of gene expression through levels of RNA species [245]. It has widely become the standard technique for validating results obtained by microarray analysis [246] and the recent commercial availability of standard fluorescent tagged probes negates the need for researchers to design multiple probes. This makes confirmation of microarray results quicker and easier. qRT-PCR does have its limitations. One of the biggest limitations of qRT-PCR for gene expression studies is the necessity to use RNA which is then reverse transcribed to cDNA. RNA isolation must be performed with the utmost care to minimise the risk of contaminants (such as genomic DNA) and maintain the integrity of the RNA itself (reviewed in [247]). Design and testing of primers must also be stringent to minimise the likelihood of non-specific binding. Compounds found within tissues may also inhibit PCR. A number of measures were used in this study to minimise error in the PCR reactions. All RNA samples were tested prior to reverse transcription to identify any loss of integrity or possible contamination (chapter 2). RNA was treated to remove genomic contamination and primers chosen which spanned exon boundaries, so should not bind to genomic DNA in the sample. Primers used were all commercially available and were assumed to have been stringently tested, however standard curves were obtained for all primers to ensure compatibility with the samples. While these measures were used to minimise error it is important to note that these limitations do exist and data should be viewed accordingly.

In this chapter I will present the results from qRT-PCR analysis of 36 genes that were identified as differentially expressed in the microarray study of the ENU mutant mice. It should be noted that this is only a subset (~10%) of the genes identified as being differentially expressed. These genes were chosen for further study based on the criteria outlines in chapter 4. I will also describe the reanalysis of 20 genes that were tested prior to outlier removal described in chapter 4, and analysis of 3 drug treatment genes which

failed to validate in the drug naïve (saline) L100P mice, but which showed persistent differential expression in the drug treated group.

5.2 Validating the Disc1 microarray result

As described in Chapter 4, prior to the discovery of the outlier group qRT-PCR had been carried out with some genes showing differential expression in the same direction seen in the microarray experiment. When the array data was re-analysed these genes would not have been selected for validation analysis (selection criteria outlined in chapter 4 section 6), and in most cases did not make the criteria for differential expression (FC +/- 1.3, $p<0.05$). None-the-less, as these genes had already been tested the data was re-analysed with the outlier pool removed, and any genes which continued to show differential expression on qRT-PCR were tested on a second batch of mouse samples, alongside the new gene list. Two batches of mouse RNA were used for qRT-PCR validation. Batch 1 was the RNA that went on the array, and batch 2 an independent sample set from a colony derived from the original animals.

5.2.1 Initial Validations and Outlier removal of the L100P drug naïve adult microarray

Twenty genes identified as differentially expressed in the comparison between the L100P drug naïve (saline) animals and C57BL/6J animals had been tested by qRT-PCR prior to the identification of the outlier pool (described in Chapter 4 section 4). When the 20 qRT-PCR data-sets were reanalysed following outlier removal, five of the 20 genes remained significant by qRT-PCR (FC +/- 1.3, $p<0.05$) (Table 5.1). *Sort1* did not exceed the fold-change cut-off imposed after the microarray, however, as this gene had previously been implicated in major mental illness it was tested in a second batch of mouse RNA alongside the five genes where the qRT-PCR result remained valid after outlier removal. Of these genes *Bex2* was also still differentially expressed at the array level. All genes tested from the initial analysis are shown in table 5.1.

Nrgn1, Litaf, Id1 and *Syn1* had significant fold-change and p-values after outlier removal, however in the opposite direction to the microarray. It is already known that between platform consistency decreases significantly with different primers or increased sequence distance [246]. These genes have multiple transcripts which arise through alternate splicing which may account for the discrepancy in the results between platforms (www.ensembl.org).

Twenty one genes that showed differential gene array expression in the L100P drug naïve saline-condition, adult mouse, post outlier removal were tested by qRT-PCR in both the RNA that was hybridised to the array, and a second batch of RNA samples. The six genes previously tested in the initial validation effort were also tested in a second batch of RNA and are displayed in table 5.1.

Gene	Array FC	Array Pval	Batch 1		Batch 2	
			Taqman Fold change	Taqman MannWhitney	Taqman Fold change	Taqman MannWhitney
Nrxn1	1.35	1.00E-03	1.34	0.041	1.29	0.0020
Wdfy1	1.46	2.85E-10	2.52	0.010	2.45	0.0022
Egr4	-1.36	5.00E-03	-2.29	0.008	-2.89	0.0210
Rapgef5	1.39	4.87E-05	1.43	0.018	1.15	0.0325
Nrxn3	1.25	2.18E-02	1.61	0.049	1.25	0.0370
Ppargc1a	1.28	1.17E-03	1.43	0.100	1.20	0.0630
Exoc4	1.26	9.25E-05	1.64	0.072	-1.10	0.1170
Gria1	1.30	4.00E-04	1.09	0.072	1.05	0.2680
s100a8	1.61	1.63E-07	1.97	0.117	1.63	0.3140
Mll5	1.29	2.46E-03	1.64	0.009	1.11	0.5000
s100a9	1.28	2.35E-05	4.45	0.015	1.59	0.5000
Pde1a	1.27	2.24E-03	1.23	0.100	-	-
Arc	-1.35	1.00E-02	-1.07	0.140	-	-
Gria2	1.27	1.00E-02	1.14	0.157	-	-
Napb	1.29	6.32E-04	1.13	0.250	-	-
Fbxo41	1.33	2.92E-03	-1.18	0.270	-	-
Dusp6	-1.28	1.80E-02	-1.01	0.300	-	-
Egr3	-1.26	1.09E-05	-1.11	0.305	-	-
Egr2	-1.26	1.00E-03	-1.31	0.377	-	-
Syn2	1.30	2.00E-03	-1.01	0.409	-	-
Cplx	1.41	1.00E-02	-1.02	0.473	-	-
Sort1	-1.05	0.3965	-1.12	0.0029	-2.72	0.0410
Shank3	-1.05	0.2343	-1.49	0.05	-2.84	0.0410
Syp	-1.11	0.0264	-1.38	0.0064	-1.39	0.1879
Ubl3	-1.06	0.0709	-1.29	0.0165	-1.07	0.2454
Gad1	-1.14	0.1262	-2.04	0.0002	-1.09	0.5000
Syn1	-1.08	0.2256	2.281	0.0005	-	-
Plk1	1.03	0.1734	19.21	0.0063	-	-
Id1	1.06	0.1517	-1.23	0.0223	-	-
Nrgn1	-1.17	0.1351	2.81	0.028	-	-
Litaf	1.06	0.2759	-2.13	0.0327	-	-
Pcbp1	1.15	0.1227	-3.02	0.0592	-	-
CamK2a	-1.12	0.0501	2.901	0.061	-	-
Gstm1	-1.08	0.1571	-1.35	0.0891	-	-
Plxnb2	1.01	0.7463	-2.86	0.1109	-	-
Dlg2	-1.09	0.0419	1.125	0.184	-	-
Gp1bb	-1.19	0.0056	1.23	0.186	-	-
Mgfe8	-1.04	0.3609	1.17	0.345	-	-
Grb10	1.01	0.7582	1.18	0.3916	-	-
Emid2	1.01	0.8448	-1.04	0.4585	-	-
Apoe	-1.04	0.5778	2.43	0.48	-	-

Table 5.1: Table of genes differentially expressed in the L100P drug naïve adult that were tested by qRT-PCR. Column 1 gives gene name, column 2 and 3 the array fold-change and p-value to be validated, columns 4 and 5 the first round qRT-PCR fold-change, and p-value for Mann-Whitney U test. Columns 6 and 7 give these values for the second round of qRT-PCR on independent RNA samples. Significant p-values are denoted in red text for both qRT-PCR rounds. Genes in the bottom half of the table below the thick dividing line were carried over from the original analysis. Genes not tested in the second round are denoted with '-' in those columns.

Four genes (*wdfy1, Nrxn1, Nrxn3 and Egr4*) from the post outlier removal data set validated robustly through both rounds of analysis. *Rapgef5* failed to meet the fold-change cut-off in the second round of qRT-PCR, even though it gave a significant p-value in statistical testing. Of the genes carried forward from the pre-outlier removal list, *Sort1* and *Shank3* both gave significant valid results in the second round of qRT-PCR analysis.

5.2.2 Validation of the Q31L drug naïve adult microarray

Only two genes were selected from the Q31L drug naïve array for validation by qRT-PCR. *Wdfy1* and *Ndn* were both up-regulated in the Q31L adult compared to C57BL/6J wild-type controls. Differential expression of both genes was validated in the first qRT-PCR analysis, however both failed to replicate when tested in a second batch of independent RNAs using wild-type littermates as controls (table 5.2).

			Batch 1		Batch 2	
Gene	Array FC	Array Pval	Taqman Fold change	Taqman MannWhitney	Taqman Fold change	Taqman MannWhitney
WDFY1	1.5	4.90E-11	2.76	0.0016	1.12	0.116
NDN	1.36	2.00E-02	1.9	0.0007	1.03	0.433

Table 5.2: Table of genes differentially expressed in the Q31L drug naïve adult that were tested by qRT-PCR. Column 1 gives gene name, column 2 and 3 the array fold-change and p-value to be validated, columns 4 and 5 the first round qRT-PCR fold-change, and p-value for Mann-Whitney U test. Columns 6 and 7 give these values for the second round of qRT-PCR on independent RNA samples. Significant p-values are denoted in red text for both qRT-PCR rounds.

A further three genes that showed correction with drug treatment were also tested in the Q31L mouse and will be discussed in section 5.2.4.

5.2.3 Validation of the L100P and Q31L embryonic microarray

Seven genes were analysed using qRT-PCR for the embryonic array samples. Three of these genes, *Pde4d, Coro1b* and *Numb* were differentially expressed in both the L100P and Q31L embryonic samples in the array analysis (table 5.3).

				Batch 1		Batch 2	
Gene	Comparison	Array FC	Array Pval	Taqman Fold change	Taqman MannWhitney	Taqman Fold change	Taqman MannWhitney
Prmt6	C57E13/L100PE13	-2.24	0.007	1.09	0.264	-	-
CORO1B Q31	C57E13/Q31LE13	-1.72	0.012	5.28	0.0002	-	-
NUMB Q31	C57E13/Q31LE13	-1.71	0.023	1.99	0.0006	-	-
NUMB L100	C57E13/L100PE13	-1.65	0.040	1.51	0.05	-	-
CORO1B L100	C57E13/L100PE13	-1.61	0.026	4.08	0.0002	-	-
PRM3	C57E13/L100PE13	-1.52	0.011	-1.35	0.058	-1.63	0.002
PDE4D	C57E13/Q31LE13	1.42	0.007	1.17	0.319	1.16	0.5
Cdh11	C57E13/L100PE13	1.55	0.049	1.81	0.002	1.45	0.002
Cstf1	C57E13/L100PE13	1.64	0.048	2.93	0.005	1.07	0.245
PDE4D	C57E13/L100PE13	1.94	0.009	2.27	0.015	1.11	0.089

Table 5.3: Table of genes differentially expressed in the L100P and Q31L embryo that were tested by qRT-PCR. Column 1 gives gene name, column 2 comparison, column 3 and 4 the array fold-change and p-value to be validated, columns 5 and 6 the first round qRT-PCR fold-change and p-value for Mann-Whitney U test. Columns 7 and 8give these values for the second round of qRT-PCR on independent RNA samples. Significant p-values are denoted in red text for both RT-qPCR rounds. Genes not tested in the second round are denoted with '-' in the appropriate columns.

Of the seven genes tested, four met the criteria for significant differential expression in qRT-PCR analysis of the RNA that was hybridised to the array. *Prm3* and *Cdh11* also validated in the second round of analysis in the L100P embryonic samples. *Pde4d* validated in the first batch in the L100P embryo but failed to replicate in the independent sample, and did not validate in the Q31L embryonic samples at either stage of analysis. *Coro1b* and *Numb* unfortunately gave significant results in the opposite direction to those previously reported in the microarray analysis in both the L100P and Q31L embryonic samples. This negated them from further analysis at this stage.

5.2.4 Validation of the L100P and Q31L drug treatment microarray

As previously mentioned, there were a number of genes that were differentially expressed in the adult samples that showed correction of expression under drug treatment conditions. Correction of expression was defined as differential gene expression (FC +/-

1.3, $p<0.05$) in the drug treated group when compared to the homozygous mutant, but not when compared to the wild-type control. Of these, eight genes were corrected in the L100P Rolipram treated group and three in the Q31L Bupropion treated group. One gene corrected by Rolipram treatment was also corrected by Clozapine treatment according to the microarray result, although this did not validate by qRT-PCR. These genes were not tested as part of the L100P adult analysis as in some cases they were slightly below the fold-change cut-off, although became interesting for analysis when compared to the drug treated individuals. These eleven genes were analysed by qRT-PCR (Table 5.4) in both treatment groups (drug naïve vs wild-type control, and drug naïve vs drug treated) through two rounds where the first batch were consistent with the microarray data.

				Batch 1		Batch 2	
Gene	Comparison	Array FC	Array Pval	Taqman Fold change	Taqman MannWhitney	Taqman Fold change	Taqman MannWhitney
Atp5b	C57BL6/L100P	1.23	0.02	1.38	0.004	1.11	0.267
Atp5b	L100P/L100P Roli	-1.37	0.04	-2.27	0.0002	-2.58	0.003
Bex2	C57BL6/L100P	-1.3	0.00004	-3.5	0.0019	1.41	0.018
Bex2	L100P/L100P Roli	1.22	0.06	2.85	0.0364	-1.99	0.0013
Dusp1	C57BL6/L100P	-1.05	0.34190	-1.26	0.026	-1.04	0.418
Dusp1	L100P/L100P Roli	1.52	0.00451	2.45	0.0006	2.56	0.008
Mast3	C57BL6/L100P	-1.19	0.001	-1.3	0.12	-1.07	0.095
Mast3	L100P/L100P Roli	1.26	0.03	1.49	0.005	2.49	0.048
Ndfip	C57BL6/L100P	1.45	0.009	1.979	0.0041	1.006	0.349
Ndfip	L100P/L100P Roli	-1.72	0.019	-3.379	0.0006	-1.51	0.009
PAK3	C57BL6/L100P	1.35	0.0002	1.03	0.0338	1.19	0.05
PAK3	L100P/L100P Roli	-1.28	0.049	-1.74	0.0023	-1.66	0.033
POP4	C57BL6/L100P	-1.62	0.005	1.09	0.026	-	-
POP4	L100P/L100P Roli	1.7	0.05	-1.1	0.007	-	-
POP4	L100P/L100P Cloz	1.73	0.05	1.02	0.42	-	-
RANBP9	C57BL6/L100P	1.14	0.07	-1.15	0.1772	-	-
RANBP9	L100P/L100P Roli	-1.33	0.04	1.31	0.0637	-	-
1500015O10RIK	C57BL6/Q31L	1.46	0.03	1.31	0.21	-	-
1500015O10RIK	Q31L/Q31L Bupro	-1.56	0.045	-1.45	0.155	-	-
GPR88	C57BL6/Q31L	1.26	0.003	2.86	0.0083	-2.56	0.05
GPR88	Q31L/Q31L Bupro	-1.32	0.03	-1.62	0.1	9.6	0.003
SYN1	C57BL6/Q31L	-1.15	0.64	2.7	<0.0001	-	-
SYN1	Q31L/Q31L Bupro	1.32	0.0004	-1.69	0.0003	-	-

Table 5.4: Table of genes differentially expressed in the L100P and Q31L drug treated adults that were tested by qRT-PCR. Column 1 gives gene name, column 2 comparison, column 3 and 4 the array fold-change and p-value to be validated, columns 5 and 6 the first round RT-qPCR fold-change, and p-value for Mann-Whitney U test. Columns 7 and 8 give these values for the second round of qRT-PCR on independent RNA samples. Significant p-values are denoted in red text for both qRT-PCR rounds. Genes not tested in the second round are denoted with '-' in the appropriate columns.

Ranbp9, *Pop4*, *1500015O10RIK* and *Syn1* gave results in the opposite direction to the array negating them from further study. More concerning was the result from *Bex2* and

Gpr88, which replicated in the first round analysis but gave significant fold-changes in the opposite direction in the second round. As this was the same platform, probe incompatibility can be ruled out in these cases suggesting there was another factor involved. qRT-PCR was repeated for both these genes in the second batch of samples to check for experimental error and the result was found to be the same in both cases.

Pak3 gave a consistent result, with higher p-values and lower fold change in the L100P drug naïve comparison across both rounds of testing. *Mast3*, *Atp5b*, *Ndfip* and *Dusp1* only replicated in the L100P drug naïve vs L100P Rolipram treated in the batch 2 analysis. To determine if this change in expression was due to genotype or drug treatment I drug treated eight C57BL/6J adult mice and ran the qRT-PCR analysis of these four genes on RNA samples from the C57BL/6J Rolipram treated and drug naïve mice (Table 5.5).

		Batch 2 - L100P vs L100P Rolipram		Batch 2 - C57BL6 vs C57BL6 Rolipram	
Gene	Comparison	Taqman Fold change	Taqman MannWhitney	Taqman Fold change	Taqman MannWhitney
Mast3	C57BL6/C57BL6 Rolipram	2.49	0.048	1.23	0.433
Atp5b	C57BL6/C57BL6 Rolipram	-2.58	0.003	1.13	0.0043
Ndfip	C57BL6/C57BL6 Rolipram	-1.51	0.009	-1.17	0.018
Dusp1	C57BL6/C57BL6 Rolipram	2.56	0.008	1.94	0.0012

Table 5.5: Table of genes differentially expressed in the C57BL/6J drug treated adults that were tested by qRT-PCR. Column 1 gives gene name, column 2 comparison, columns 3 and 4 the second round qRT-PCR fold-change, and p-value for Mann-Whitney U test from the L100P drug naïve vs L100P Rolipram treatment. Columns 5 and 6 give these values for the second round of qRT-PCR on independent RNA samples from C57BL/6J mice treated with the same rolipram dose compared to drug naïve C57BL/6J mice. Significant p-values are denoted in red text for both qRT-PCR rounds.

Mast3 did not show significant differential expression in the C57BL/6J drug treated mice suggesting a possible effect of genotype on drug action. *Atp5b*, *Ndfip* and *Dusp1* all gave significant results in the C57BL/6J Rolipram treated group which would suggest that the results seen in the L100P Rolipram treated mice were more a factor of drug treatment than genotype. However, the *Atp5b* differential expression in the C57BL/6J comparison is in the opposite direction to the L100P comparison, which may suggest some effect of

genotype on the drug action. This would require further study to determine if this is the case.

5.3 Conclusions

Of the 40 genes tested by qRT-PCR only seven gave robust and reliable results through both rounds of validation. Of these, four were from the L100P drug naïve adult group (*Wdfy1, Nrxn1, Nrxn3 and Egr4*), two from the L100P embryonic group (*Cdh11, Prm3*) and one from the L100P Rolipram treated group which showed correction (*Pak3*). One *(Rapgef5*) gave p-value<0.05 but did not reach the 1.3 fold-change cut-off implemented in earlier analysis. Of the five genes carried over from the pre-outlier removal gene list, two validated through both rounds of RT-qPCR with significant fold-change and p-value to meet the criteria imposed. While microarrays are powerful molecular biological tools, they have technical aspects which can produce results that erroneously under or over-represent specific genes. False negativity can result from low expression levels, inefficient priming of specific mRNAs resulting in transcript drop out, poor adhesion of DNA to the slide, and splice variants with sequences not included on the array. Conversely, sources of false positivity include repetitive nucleotide elements, sequence homology between functionally different transcripts and high background levels due to nonspecific binding of nucleotides [248]. This not only justifies the need to validate results through other platforms, but may help to explain why two genes that were not seen to be differentially expressed on the array platform post outlier removal, may be seen to be differentially expressed on the qRT-PCR platform.

The neurexin family of proteins function as cell adhesion molecules and receptors. Neurexins 1-3 utilize two alternate promoters and include numerous alternatively spliced exons to generate thousands of distinct mRNA transcripts and protein isoforms. The majority of transcripts are produced from the upstream promoter and encode alpha-neurexin isoforms; a much smaller number of transcripts are produced from the downstream promoter and encode beta-neurexin isoforms. The numerous isoforms of

Neurexin 1 (*NRXN1*) i sinvolved in maintenance of synaptic junctions and mediate intracellular signalling[249]. By making heterophillic connections with neuroligands it leads to maturation and differentiation of GABAergic and glutamatergic synapses through bi-directional signalling [250]. *NRXN1* was implicated by Walsh *et al* (2008) [217] in an analysis of copy number variants in schizophrenia. Other studies on copy number variants have also implicated *NRXN1* in schizophrenia pathology. Need *et al* (2009) [251] found deletions in the 3' end of *NRXN1* in patients with schizophrenia, but not in control groups. Data from Vrijenhoek *et al* (2008) [252] suggest that CNVs that affect the first few exons of *NRXN1* confer greater risk of major mental illness, while that from Rujescu et al (2009) [253] implies that deletions that affect exons directly increase susceptibility. In addition, a missense mutation in exon 1 of human *NRXN1* has also been linked to autism [254]. Neurexin 3 (NRXN3) is a membrane protein involved in cell adhesion, synaptic transmission and neurotransmitter secretion [249]. SNPs in *NRXN3* have been associated with alcohol and nicotine dependence and linked with opiate dependence [255, 256].

The Early Growth Response Factors (EGRs) 2, 3 and 4 are synaptic activity inducible immediate early genes and all show nominal association with schizophrenia in a Japanese population. They activate transcription of genes required for mitogenesis and differentiation. *EGR1, 2* and *3* are downregulated in the prefrontal cortex of brains from patients with schizophrenia [238] , *EGR4* expression was too low to measure disregulation. Treatment with the atypical antipsychotic and antidepressant Aripiprazole increases expression of *Egr4* in rat frontal cortex[257].

Little is known about the function of WD repeat and FYVE domain-containing protein 1 (*WDFY1* or *FENS1*) but it is known to be ribosomal and has been associated with alcohol consumption and preference [258]. Primary functional analysis identifies *Wdfy1* as being involved in endosome trafficking [259].

Eight genes from the initial analysis showed correction of the L100P non-drug treated disregulation. Only one of these validated through qRT-PCR. Serine/threonine-protein

kinase 3 (PAK3), is involved in the Erb signalling pathway, axon guidance and focal adhesion. PAK proteins form an activated complex with GTP-bound RAS-like (P21), CDC2 and RAC1 proteins which then catalyzes a variety of targets. It is thought this protein may be necessary for dendritic development and for the rapid cytoskeletal reorganization in dendritic spines associated with synaptic plasticity. Mutations in *PAK3* have been associated with X-linked nonsyndromic mental retardation [260-262] and suppression of *PAK3* results in formation of abnormally elongated dendritic spines and a reduction of mature synapses [263]. These elongated spines fail to express post-synaptic densities or contact presynaptic terminals resulting in reduced spontaneous activity and defective LTP. *Pak3* knockout mice have deficiencies in learning and memory and abnormalities in synaptic plasticity along with a reduction of transcription factor cAMP-responsive element-binding protein (CREB) suggesting a novel signalling mechanism with PAK3 and Rho signalling regulating synaptic function and cognition [264]. *PAK* genes also interact with the known Disc1 interactor Kalirin-7 which is thought to control multiple aspects of spinal plasticity [265].

Cadherin 11 (Cdh11) is a type 2 classical cadherin involved in mediation of calcium dependant cell adhesion. Type II (atypical) cadherins are defined based on their lack of a HAV cell adhesion recognition sequence, that is specific to type I cadherins. It comprises of a large N-terminal domain (extracellular), a single membrane spanning domain, and a small intracellular C-terminal domain. It was first discovered in rodent brain samples [266] with high expression in the developing olfactory system at E13-E17 [267] and has since been linked to multiple cancer types [268, 269].

The remaining three genes that have validated in the rolipram treated L100P adults but have not validated in the L100P adult are *Atp5b*, *Ndfip* and *Dusp1*. Two of these genes (*Atp5b* and *Dusp1*) are involved in phosphate metabolism and were classed as over-enriched in the sample set by GOtree analysis. This could be interesting on two counts; the action of Rolipram may target phosphate metabolism genes, which may or may not be significant in schizophrenia. Or phosphate metabolism is a key pathway in schizophrenia. When the list of genes differentially expressed in only the L100P adult group was run

through Gotree analysis the only over enriched genes were involved in synaptic plasticity, not phosphate metabolism, suggesting this is more a function of the drug treatment than the mouse phenotype. To confirm this I tested these genes in a C57BL6 drug naïve group vs C57BL6 rolipram treated group as this would allow me to determine if the effect was due to the drug treatment or the mutation. *Dusp1* was significantly up regulated in the C57BL6 rolipram treated group compared to the drug naïve group, as it had been in the L100P rolipram treated group compared to the L100P drug naïve group. *Ndfip* also behaved in the same way in the C57BL6 rolipram treated group as it had in the L100P drug treated group, being significantly down regulated compared to the drug naïve control. *Atp5b* was significantly down regulated in the C57BL6 rolipram treated group where it had been up regulated in the L100P rolipram treated group, but a significant difference in expression was observed suggesting it is an effect of the drug treatment as opposed to the mutation. As the changes in expression in these three genes appear to be an effect of the rolipram treatment rather than the mutation I feel they do not fit the criteria for further analysis within this project.

Finally *Sortilin (Sort1)* and *Shank3* which were carried over from the original gene list, validated in the second sample set. *Sortilin* (alias neurotensin receptor 3) has been implicated in the modulation of dopamine signals [270] and was found to be downregulated in a Scottish family with a history of bipolar disorder (Christoforou et al, manuscript in preparation). *Shank3* is a structural post-synaptic density protein which has previously been implicated in mental retardation [271]. Mutation of a single *Shank3* allele has also been found in Autism spectrum disorders [272].

As described previously, some genes tested had fold-changes in opposite directions across the two platforms. This has been documented previously [246] and is thought to be due to the platform probes amplifying different isoforms of the gene in question. While every effort was made to ensure overlap of the two probes, a perfect match could not be ascertained. In all cases both probes were mapped to the same exon but as the sequences of the qRT-PCR probes are not made available identical sequence alignment could not be achieved. A more concerning observation was that *Bex2* and *Gpr88*, both genes that

displayed correction through drug treatment, gave opposite fold-change directions between sample batches. As both batches were run on the same platform, this cannot be attributed to probe amplification discrepancy. Differences in the wild-type expression could be due to minor genetic differences between the C57BL/6J and the L100P wild-type. Although the L100P animals had been backcrossed for over 10 generations it is unlikely, but not impossible, that there were still some genetic differences.

The difference in the drug treated groups is more difficult to explain. It could be speculated that differences in environmental conditions (such as feed, housing etc) can affect the effect of pharmacologicals. As the batches of RNA were collected from animals bred in different facilities (though of the same background strain) this may account for the differences seen here. Also, the animals used in the first batch underwent behavioural testing just prior to being culled. It may be that this would have an effect on gene expression and thus drug efficacy.

Additional research would ideally focus on testing the other dysregulated genes present from the microarray study, which was outwith the scope of this project. As Disc1 has been shown to be developmentally regulated [134], establishing a developmental profile for both the *Disc1* mutant, and the genes dysregulated by the mutation, would be an advantageous addition to the current data and is described in chapter 6. Protein analysis of those genes that showed robust disregulation is a logical next step for this project and is described in chapter 7.

Chapter 6
Establishing a developmental profile for the genes of interest in L100P mutant mice and wild type controls

6. Establishing a developmental profile for the genes of interest in L100P mutant mice and wild type controls

Since the 1940's and 50's scientists have considered the possibility that schizophrenia could be a developmental disease. In 1986 Weinberger proposed the developmental hypothesis of schizophrenia which suggested that small neurodevelopmental deficits in key circuits could lie dormant until puberty, when normal molecular changes could facilitate the onset of disease in affected individuals [189]. With this in mind I decided to establish a developmental profile of each of the validated genes of interest in the L100P mouse model. As Disc1 has been shown to be developmentally regulated [134] the aim was to determine if a mutation to *Disc1* would alter the developmental profile and/or the genes found to be dysregulated in the mutants by the microarray analysis.

6.1 The Disc1 developmental profile

While collecting samples for the *Disc1* microarray I had also collected a developmental profile panel from both the L100P and Q31L mice, to determine whether *Disc1* expression was altered during development. A Disc1 developmental profile was previously published by Shurov et al (2004) [134] describing peaks in Disc1 protein expression at E13 and P35, corresponding to periods of neurogenesis in the developing whole brain and puberty in the adult mouse . The study used cortex from the adult mouse and whole brain from mouse embryos to determine protein expression through western blotting. However, Shurov *et al* did not specify the mouse strain used in their study. As each inbred strain is genetically different, I decided to establish a developmental profile for both ENU lines and wild-type counterparts to be used in this study, to determine whether the mutations affected the expression of *Disc1* through development.

Mice were taken at eight stages throughout development from both ENU mutant lines, their wild-type littermates and C57BL/6J animals which were used as controls in the microarray experiment. Whole brain was used for embryonic samples and hippocampus from mature mice matching the stages used in the array experiment. There were six mice at each stage, for each line, giving a total of 240 samples to be run in triplicate. I used a

"pan" Disc1 probe, designed to hybridise at exons 4 and 5, thus covering all Disc1 isoforms bar one (the extreme short isoform Es: contains an alternatively spliced DISC1 exon 1a and terminates transcription two intronic codons after exon 3) to establish a developmental profile in these animals (Figure 6.1).

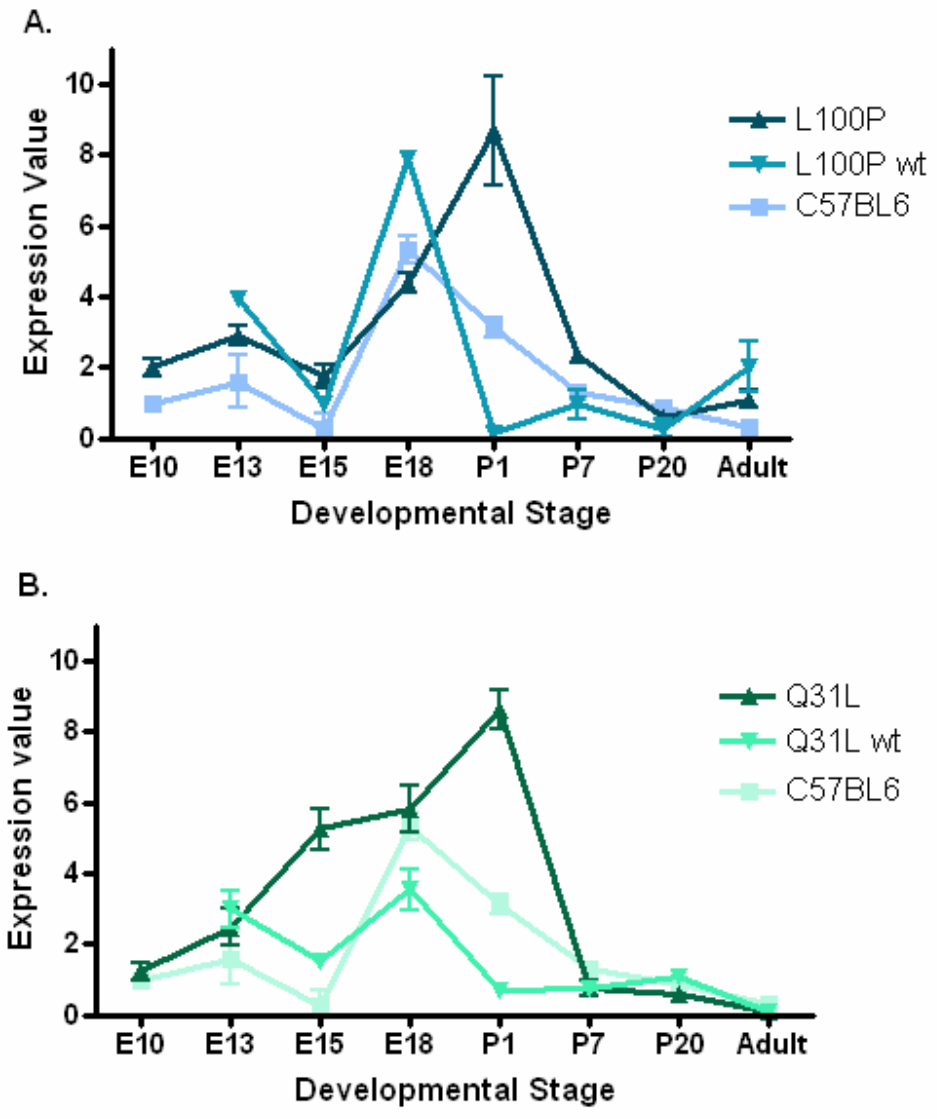

Figure 6.1 Developmental Profile of Disc1 in the mouse brain. Graph A shows the expression profile of Disc1 in the brain of L100P wild type littermates, L100P mutant and C57BL/6J mouse lines through 8 stages of development. Graph B show the expression profile of Disc1 in the brain of Q31L wild type littermates, Q31L mutant and C57BL/6J mouse lines through 8 stages of development.

I performed a one-way analysis of variance with a post-hoc Tuckey test to determine where the strains differed significantly in *Disc1* expression across the panel tested. There was a significant difference across the whole dataset ($F=41.65$, $df=34$, $p<0.0001$) with within stage significant differences arising at E15, E18 and P1.

With regard to *Disc1* expression, the homozygous L100P mutant is significantly different from the L100P wild type littermate and the C57BL/6J ($p<0.001$) at E13. At E18, the L100P mutant and the L100P wild-type are not significantly different than each other, but are significantly different ($p<0.001$) to the C57BL/6J. At P1, the L100P mutant has significantly higher expression than both the C57BL/6J and the L100P wild-type controls. As reported previously (Clapcote *et al* 2007) [197], there was no difference in *Disc1* expression levels in the adult mouse.

The homozygous Q31L mutant has significantly higher *Disc1* expression ($p<0.001$) than the C57BL/6J animals. At E18, the Q31L wild-type differs significantly from the Q31L homozygous mutant and the C57BL/6J ($p<0.05$). Perhaps the greatest difference was observed at P1, where the C57BL6 mouse is significantly different to all other groups ($p<0.001$) and the Q31L animals have significantly higher expression than their wild-type controls.

It can be speculated that the different developmental expression patterns of the Q31L and L100P mutant may contribute to the different phenotypes observed in behavioural testing. This result gives rise to two important questions; what neurobiological processes are occurring at these stages where Disc1 expression is altered and was the C57BL6 a suitable control animal for the microarray experiment given the results?

A review of the literature on mouse developmental stages gave very few results but one review article can shed some light on the processes involved at stages E15-P1 where the Disc1 expression differences have been observed. Matsuki et al (2005) [223] carried out a microarray experiment to determine which clusters of genes were most highly expressed at stages E12, E15, E18 and P1. Their data suggests that, based on genes with known

neural functions, the functional gene clusters most highly expressed at E15 are involved in neural determination and differentiation, and at E18 and P1 are involved in synapse formation and function, and survival and growth. Based on all genes differentially expressed at E18 and P1 the top functions are intra/inter cellular molecular transport, and signal transduction.

From the results obtained from the developmental panel I would suggest that the C57BL/6J mouse was an adequate control at the stages (E13 and adult) used in this study as their expression of *Disc1* did not differ significantly from the wild-type littermates at these stages. However, the results suggest that either the ENU lines were not suitably backcrossed for the C57BL/6J to be an ideal control, or there was another mutation present in the ENU mice which was retained throughout the backcrosses in the 'wild-type' control animals. For this reason the remainder of the developmental panels were run with both 'control' groups for reference, although only the L100P wild-type animals would be used for comparisons.

6.2 Establishing developmental profiles for the genes of interest

As all genes that were successfully validated came from the L100P comparisons, the genes were only tested in the L100P developmental panel. Due to constraints on time and finances I pooled the samples at each developmental stage resulting in a total of 24 samples to be run in triplicate, compared to 144 if left unpooled.

6.2.1 The Disc1 developmental profile on pooled samples

As I had pooled the samples for subsequent gene panels I decided to run the *Disc1* panel on pooled samples to check it was comparable to the previous result. Figure 6.2 shows the pooled samples compared to unpooled samples across all stages.

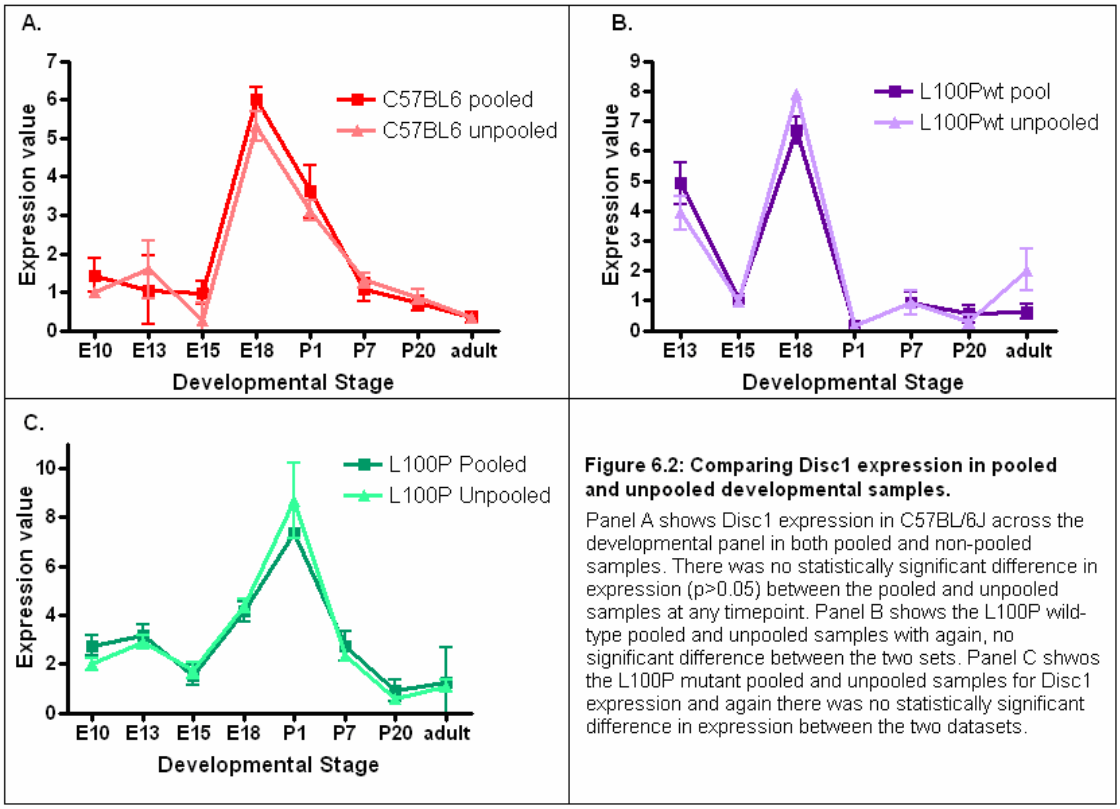

Figure 6.2: Comparing Disc1 expression in pooled and unpooled developmental samples.

Panel A shows Disc1 expression in C57BL/6J across the developmental panel in both pooled and non-pooled samples. There was no statistically significant difference in expression ($p>0.05$) between the pooled and unpooled samples at any timepoint. Panel B shows the L100P wild-type pooled and unpooled samples with again, no significant difference between the two sets. Panel C shwos the L100P mutant pooled and unpooled samples for Disc1 expression and again there was no statistically significant difference in expression between the two datasets.

While absolute expression values in the pooled and unpooled samples were not identical, the trend of expression was the same and there was no statistically significant difference between the datasets ($p>0.05$) suggesting little effect of pooling. As a new dilution of calibrator DNA was used for the pooled samples, the absolute expression values are graphed as normalized to both calibrator samples. These data indicate that pooled samples should give comparable data to the individual samples for subsequent developmental panels.

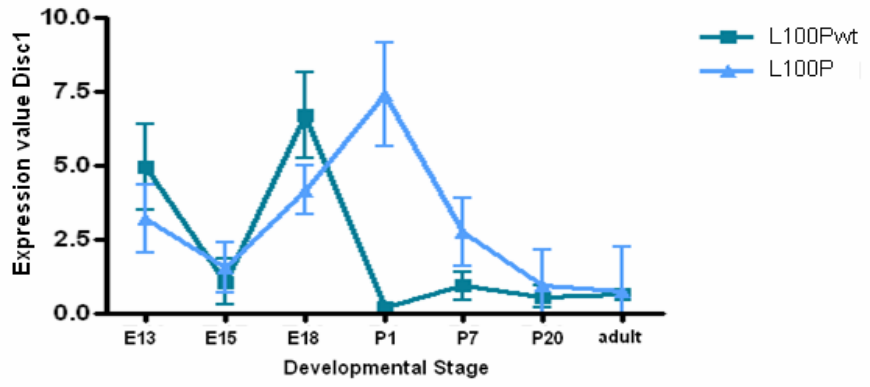

Figure 6.3: Developmental profile of Disc1 in the L100P mutant mouse. Graph shows the comparable Disc1 developmental profiles in the brains of pools from the L100P wild type and L100P mutant mouse. There is a notable time shift in peak expression in the L100P mutant mouse.

When comparing the L100P mutant and wild-type littermates over the developmental panel, what is perhaps most evident, is the appearance of a shift in the peak of *Disc1* expression during development. The L100P wild type expression values peak at E18 and the drop to level out after birth. The L100P mutant however does not show a peak in expression until P1, with a shallower decline to adulthood, with the two lines converging by P20. This delay in the L100P mutant mouse could have detrimental effects on neurodevelopment in the mutant animals. Personal observations and those of other researchers working with these animals show that the mutant individuals are physically smaller when born, which could signify a delayed development (Clapcote *et al*. personal communication). While the result of one gene expression profile is not enough to stipulate cause and affect it is worth noting that low birth weights have previously been associated with schizophrenia [273] and a latent shift in the expression of key developmental proteins could be a factor in this. Hikida *et al* [193] also note that there is a 3month delay in neuronal migration in a *Disc1* mouse model, which could potentially be attributed to a change in *Disc1* expression during development. While this chapter will only describe the developmental profile results of genes of interest from the microarray study I believe this altered *Disc1* expression during development is worthy of further investigation in the future.

6.2.2 Developmental Profile for *Egr4*

As mentioned in chapter 5, *Egr4* was the gene with the largest fold change in the sample set and a good gene for further analysis due to its nominal association with schizophrenia in a Japanese association study [238].

As with the pooled *Disc1* panel, the *Egr4* probe was run against a pooled DNA panel across seven developmental stages from E13 to adult (figure 6.4). The *Disc1* panel is being shown for comparison.

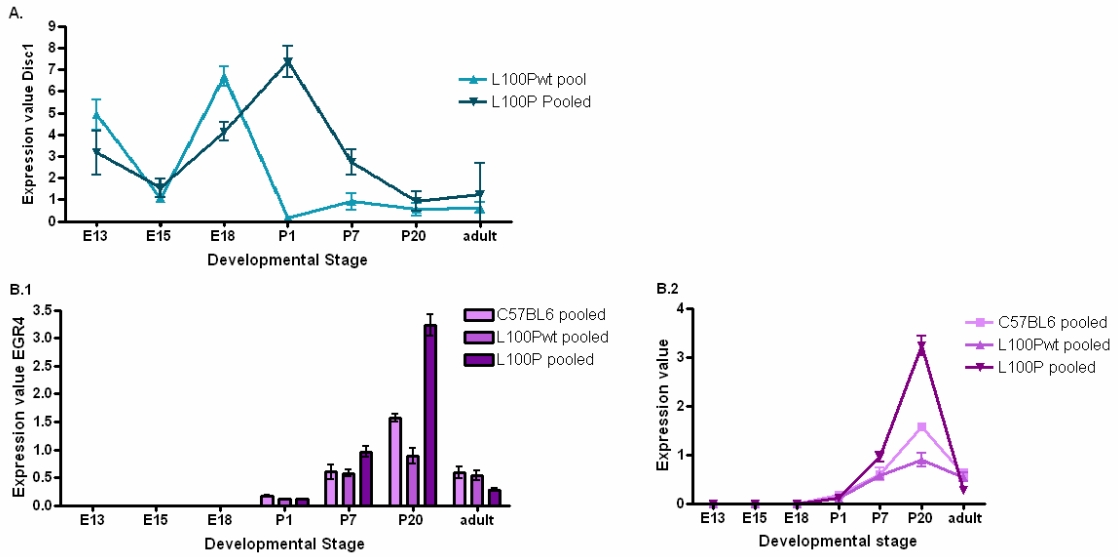

Figure 6.4: Developmental profile of EGR4 in cDNA pools from the L100P mutant mouse. Graph A shows the developmental expression patterns of Disc1 in the brain of both the L100P wild-type and L100P mutant mouse. Graph B.1 shows the developmental expression pattern of EGR4 in the brain of the L100P wild-type, L100P mutant and C57BL/6J mouse as a bar chart and B.2 as a line graph. In all cases the X axis shows the developmental stage, and the Y axis the arbitrary expression value from Taqman RT-qPCR analysis.

The Taqman RT-qPCR analysis could not detect *Egr4* in stages E13-E18. Levels of expression in all lines peaked at P20 with significant differences in expression between the L100P wild-type and L100P mutant samples at P7 ($p=0.029$) and P20 ($p=0.0007$), as well as the adult stage which was reported in chapter 5 ($p=0.0465$).

6.2.3 Developmental Profile for *Nrxn1* and *Nrxn3*

As mentioned previously, NRXN1 is involved in maintenance of synaptic junctions and is a mediator of intracellular signalling. A missense mutation in exon 1 of human *NRXN1* has been linked to autism [254] and many studies have noted CNV's in *NRXN1* and *NRXN3* linked to schizophrenia [83, 217, 251-253]. NRXN3 is a membrane protein involved in cell adhesion, synaptic transmission and neurotransmitter secretion. SNPs in *NRXN3* have been associated with alcohol and nicotine dependence and linked with opiate dependence [255, 256]. The developmental panel for both *Nrxn1* and *Nrxn3* are shown in Figure 6.5.

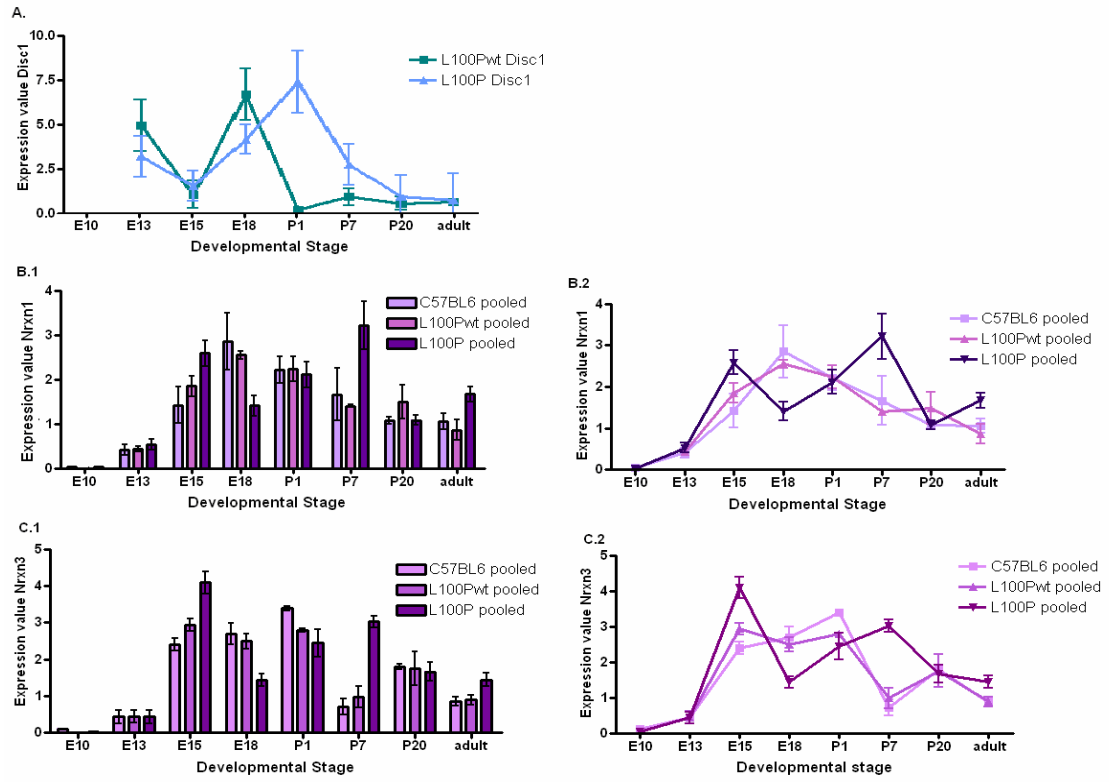

Figure 6.5: Developmental profile of Nrxn1 and Nrxn3 in cDNA pools from the L100P mutant mouse. Graph A shows the developmental expression patterns of Disc1 in the brain of both the L100P wild-type and L100P mutant mouse. Graph B.1 show the developmental expression pattern of Nrxn1 in the brain of the L100P wild-type, L100P mutant and C57BL/6J mouse as a bar chart and B.2 as a line graph. Graph C.1 shows the developmental expression pattern of Nrxn3 in the brain of the L100P wild-type, L100P mutant and C57BL/6J mouse as a bar graph and C.2 as a line graph. In all cases the X axis shows the developmental stage, and the Y axis the arbitrary expression value from Taqman RT-qPCR analysis.

The most obvious observation from both panels is the inverted trend of expression from E15 to P1 in both lines. That is, the L100P wild-type and L100P mutant show almost exact opposites in trend of expression at these stages. For *Nrxn1* there is a peak in expression at E18 in the wild-type lines, while this stage represents a dip in expression in the mutant lines (*Nrxn1* p= 0.01). *Nrxn1* expression in the L100P mutant peaks at P7 with significantly different expression to the wild-types (p=0.029) and shows significant over-expression in adulthood (p=0.048), confirming the previously reported result from chapter 5. *Nrxn3* is significantly overexpressed at E15 (p=0.026), P7 (p=0.0036) and adulthood (p=0.04) and is significantly lower in the L100P mutants at E18 (p=0.016). The mutant lines show two peaks, at E15 and P7. At E15 neurogenesis occurring in the developing mouse brain and at P7 neurite outgrowth, myelination, synaptic pruning and apoptosis occur [274]. Given the later peak of *Disc1* expression reported in the L100P mutants, and lower birth weights observed, it is possible there is a developmental delay in synaptogenesis in these animals but this cannot be confirmed without further physiological examination. As the predicted developmental function at E18 is the formation of synapses, the dip in expression of genes involved in synaptic maintenance and transmission in the mutant animals is also of key importance.

6.2.4 Developmental Profile of *Pak3*

As mentioned in chapter 5, Pak3 is involved in the Erb signalling pathway, axon guidance and focal adhesion. Mutations in *PAK3* have been associated with X-linked nonsyndromic mental retardation [260, 263]. *Pak3* knockout mice have deficiencies in learning and memory, and abnormalities in synaptic plasticity. They also show reduction of transcription factor cAMP-responsive element-binding protein suggesting a novel signalling mechanism with PAK3 and Rho, regulating synaptic function and cognition [264]. *PAK* genes also interact with the known DISC1 interactor Kalirin-7 which is thought to control multiple aspects of synaptic plasticity [265]. It is also the only drug corrected gene which gave robust and consistent results through all stages of validation.

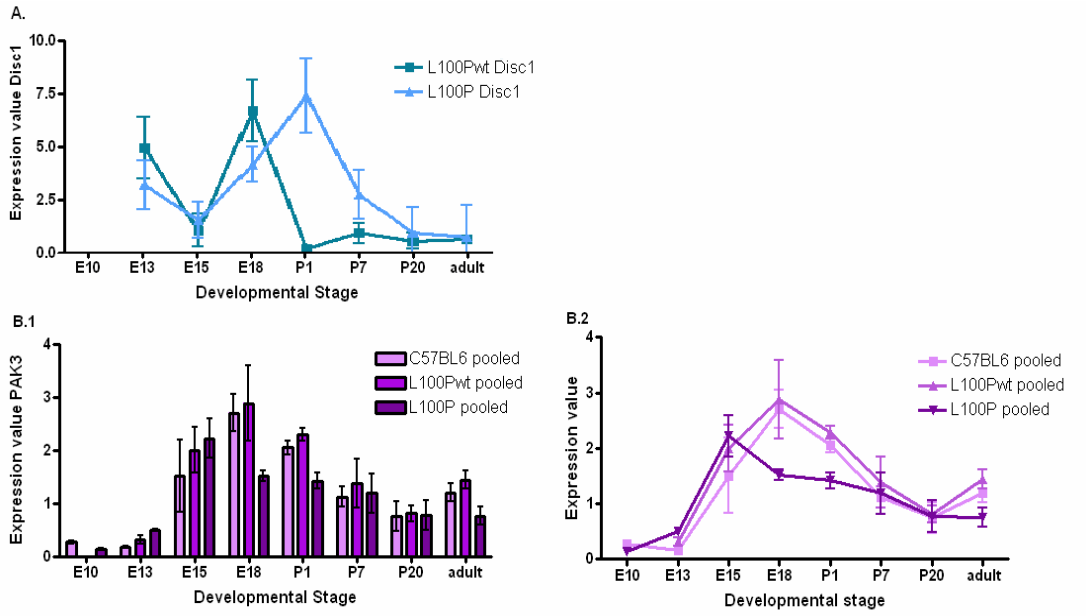

Figure 6.6: Developmental profile of *Pak3* in cDNA pools from the L100P mutant mouse. Graph A shows the developmental expression patterns of Disc1 in the brain of both the L100P wild-type and L100P mutant mouse. Graph B.1 show the developmental expression pattern of PAK3 in the brain of the L100P wild-type, L100P mutant and C57BL/6J mouse as a bar chart and graph B.2 as a line graph. In all cases the X axis shows the developmental stage, and the Y axis the arbitrary expression value from Taqman RT-qPCR analysis.

Expression of *Pak3* in the L100P wild type littermate rises from E13 to E18. At E18 there is a significant difference between the expression levels of the L100P wild type and the L100P mutant ($p = 0.049$) as expression in the L100P mutant line starts to decline after E15. There is also a significant difference in expression between the wild-type and the mutants at P1 ($p=0.0093$). The L100P mutant peaks at E15 and falls away steadily to level out at P20 to adulthood. This is the only profile where the peak in expression of the mutants precedes that of the wild-type, although the level of the peak in the mutants is considerably lower than that of the wild-type, and no difference in expression value is observed at the E15 developmental stage. The previously reported over-expression of *Pak3* in the adult mutant mice is not replicated here.

6.2.5 Developmental Profile of Cdh11

Cdh11 is a type II classical cadherin that mediates calcium dependant cell-cell adhesion. It was originally found to be differentially expressed in the embryonic samples run on the microarray. Enrichment of Cdh11 in future subplate neurons [275] has been linked to neuron extension and guidance in the developing mouse brain.

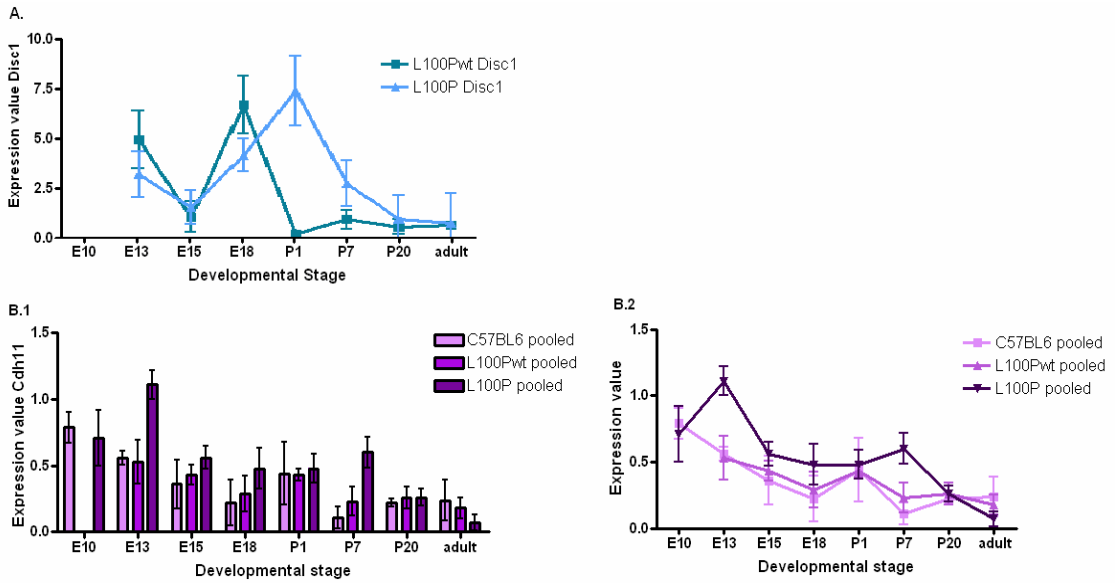

Figure 6.7: Developmental profile of *Cdh11* in cDNA pools from the L100P mutant mouse. Graph A shows the developmental expression patterns of Disc1 in the brain of both the L100P wild-type and L100P mutant mouse. Graph B.1 show the developmental expression pattern of Cdh11 in the brain of the L100P wild-type, L100P mutant and C57BL/6J mouse as a bar chart and B.2 as a line graph. In all cases the X axis shows the developmental stage, and the Y axis the arbitrary expression value from Taqman RT-qPCR analysis.

Throughout the developmental panel there is no significant difference in *Cdh11* expression between the mouse lines, except at E13 as reported in chapter 5. At E13 the expression of *Cdh11* is significantly higher (p=0.044) in the L100P mutant than in the wild type littermate. An increase in expression levels of *Cdh11* at E13 may suggest excessive neurite extension in the developing brain of the L100P mutant mouse which would affect the positioning and effectiveness of neural connections in the developed animal. While the expression profile indicates enrichment at P7 this was not statistically significant (p=0.08).

6.2.6 Developmental Profile of *Sort1*

Sort1 is one of the genes which was discarded after the identification of the outlier mouse during the microarray analysis. It did, however, validate as being dysregulated during Taqman RT-qPCR analysis (both pre and post outlier removal). For this reason, and because it was a gene of interest because it is dysregulated in a Scottish family with a history of mental illness, the developmental panel was run for *Sort1* also (figure 6.8).

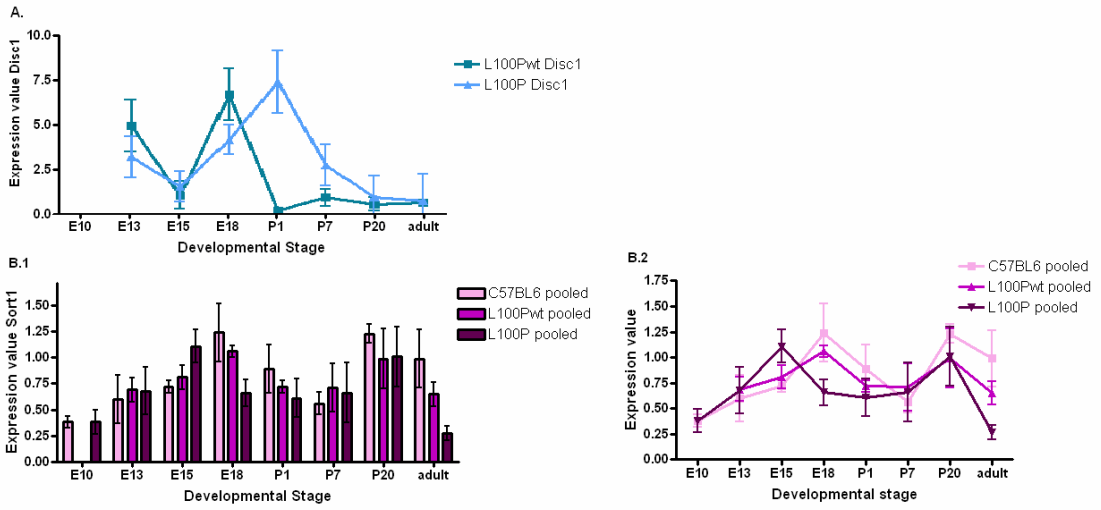

Figure 6.8: Developmental profile of *Sort1* in cDNA pools from the L100P mutant mouse. Graph A shows the developmental expression patterns of Disc1 in the brain of both the L100P wild-type and L100P mutant mouse. Graph B.1 show the developmental expression pattern of Sort1 in the brain of the L100P wild-type, L100P mutant and C57BL/6J mouse as a bar chart and B.2 as a line graph. In all cases the X axis shows the developmental stage, and the Y axis the arbitrary expression value from Taqman RT-qPCR analysis.

The L100P wild type showed peaks in *Sort1* expression at E18 and P20 with the L100P mutant showing peaks at E15 and P20. Statistical analysis of the expression levels at each time point gave significant results at E18 ($p=0.046$) and adulthood ($p=0.048$). In both cases the expression of *Sort1* was reduced in the L100P mutant animals.

6.3 Conclusions from the developmental profiles

The developmental profile of *Disc1* highlighted some key stages of development where expression is altered in the mutant mouse line, and also confirmed the previous result that expression was not significantly different at either E13 or adulthood between the L100P mutant and wild type controls. The disparity between the developmental profile of *Disc1* described in this chapter and that previously published by Shurov [134] highlights the importance of strain differences during development, and had this experiment been carried out prior to the microarray study it is likely a different developmental stage would have been chosen for the whole genome gene expression experiment.

The time shift observed in the peak of *Disc1* expression could signify a delay in neural development of the L100P mutant, although this cannot be confirmed without further physiological analysis. Further weight is added to this suggestion by a similar time shift in the peak expression of both the *neurexin* genes and *Pak3* (summary figure 6.9).

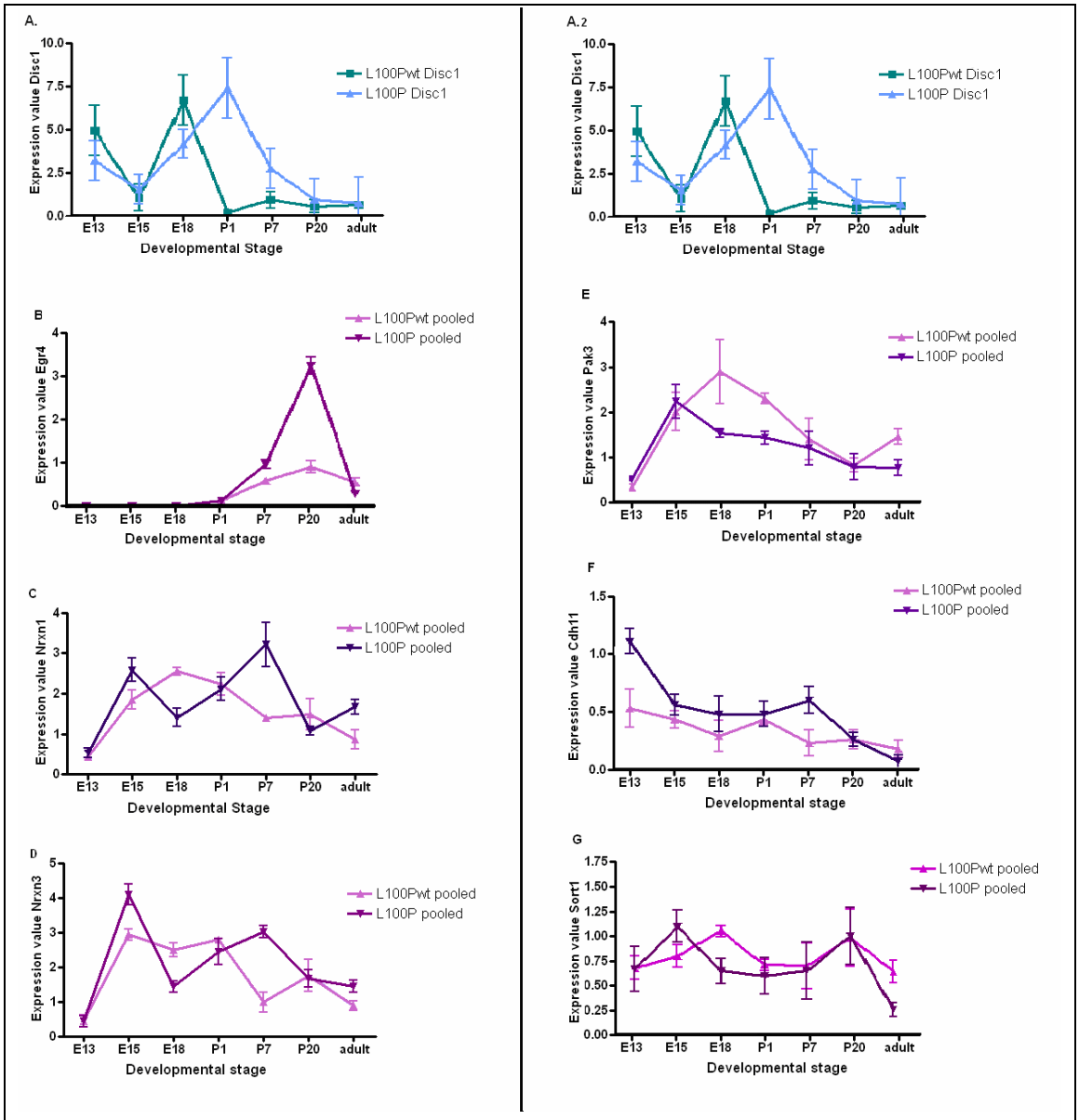

Figure 6.9: Composite of the developmental profiles compared to Disc1. A and A2 are both the Disc1 developmental panel with B:Egr4, C:Nrxn1, D:Nrxn3, E:Pak3, F:Cdh11 and G:Sort1.

The lack of definable expression of *Egr4* in the embryonic stages is of great interest and has not been previously reported. Its significantly different levels of expression at P20, just prior to when the animal is effectively going through puberty, is also of interest due to the significance of pubescent molecular changes in the brain in the developmental hypothesis of schizophrenia [189]. If this is the case then the altered expression of the

other genes studied during development could be responsible for deficits in key circuitry which are exposed at this stage facilitating the disease phenotype, as described in Weinbergers developmental hypothesis [189], in the L100P mutant model. It is my opinion that these findings warrant further investigation into the developmental process of the schizophrenia-like phenotype in this mouse model, which may shed light on the development of the human disease.

Chapter 7
Preliminary protein analysis of validated genes in L100P drug naïve (saline) adult mouse samples

7. Preliminary protein analysis of validated genes in L100P drug naïve adult mouse samples

In this chapter I will present the results from semi-quantitative western blotting and immunocytochemistry for five of the genes validated through qRT-PCR in the L100P adult mouse mutant. *Wdfy1* and *Cdh11* were not tested at the protein level as suitable antibodies were not available.

Validation of the changes at the protein level allows the investigation of alterations in isoform production and assessment of possible alterations in expression patterns. The data described here is not a full analysis but rather a first step in protein analysis for the genes identified in the array.

7.1 Protein analysis of Egr4

Egr4 had previously validated in an independent RNA sample set with a fold-change of -2.89 and a p-value of 0.013. Protein extracted from the same animals used to collect the independent RNA sample was used for western blotting. Hippocampal neurons cultured from a third batch of adult mice were used for immunocytochemistry after 21 days in culture (figure 7.1). Egr4 is expressed as a 50 kDa protein. Three replicates of each western blot was carried out and the results quantified using the Image J programme, which takes a scanned image of the blot and measures the intensity of each band. Two replicates and the quantification analysis were carried out by myself, and one replicate by Susan Anderson.

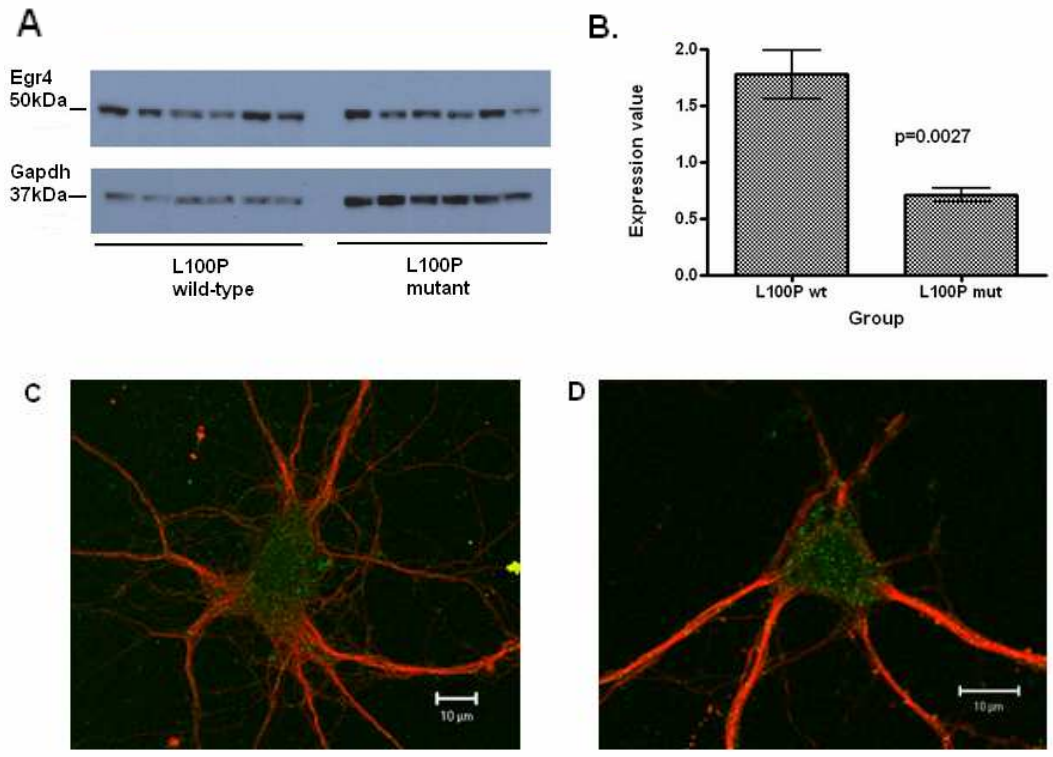

Figure 6.1: Western blotting and ICC for EGR4 in L100P ENU mutant mice. Panel A shows a representative western blot for Egr4 with the Gapdh membrane used for correction of loading differences below. Panel B shows the semi-quantative analysis of the blot using image J and a standard student t-test ($p<0.01$). Panel C shows the confocal microscope image of wild-type primary cultured hippocampal neurons (day 21) stained for Egr4 (green) and co-stained for β-actin (red). Panel D shows the mutant primary cultured hippocampal neurons stained for Egr4 (green) and co-stained for β-actin (red).

As can been seen from the figure, the disregulation of *Egr4* previously described at the RNA level is maintained at the protein level. There appears to be no overall change in expression patterns of Egr4 in the cultured neuron between wild-type and mutant animals. In both cases there are few disperse puncta on the dendrites, with the majority of Egr4 expression contained within the soma. Disc1 has been shown to be expressed in similar puncta in the dendritic shaft with localisation to the ribosomes and microtubule structures in the soma [276]. It can be speculated that Egr4 has a similar, although not identical, expression pattern to Disc1 in the neuron, although further staining of specific cell components would be required to confirm this. Quantitative analysis of the neuron

cultures was not carried out so small changes not obvious to the naked eye cannot be ruled out, but this would require further analysis, which will be discussed later.

7.2 Protein analysis of Pak3

The mental retardation gene *Pak3* was found to be differentially expressed in the L100P drug naïve adults in the microarray analysis, and was corrected to a level not significantly different to wild type when the animals were treated with the PDE4 inhibitor rolipram. This differential expression was confirmed in two sample batches by qRT-PCR. Pak3 is expressed as a 75kDa protein (figure 7.2). Again, one of the western blot replicates was carried out by Susan Anderson.

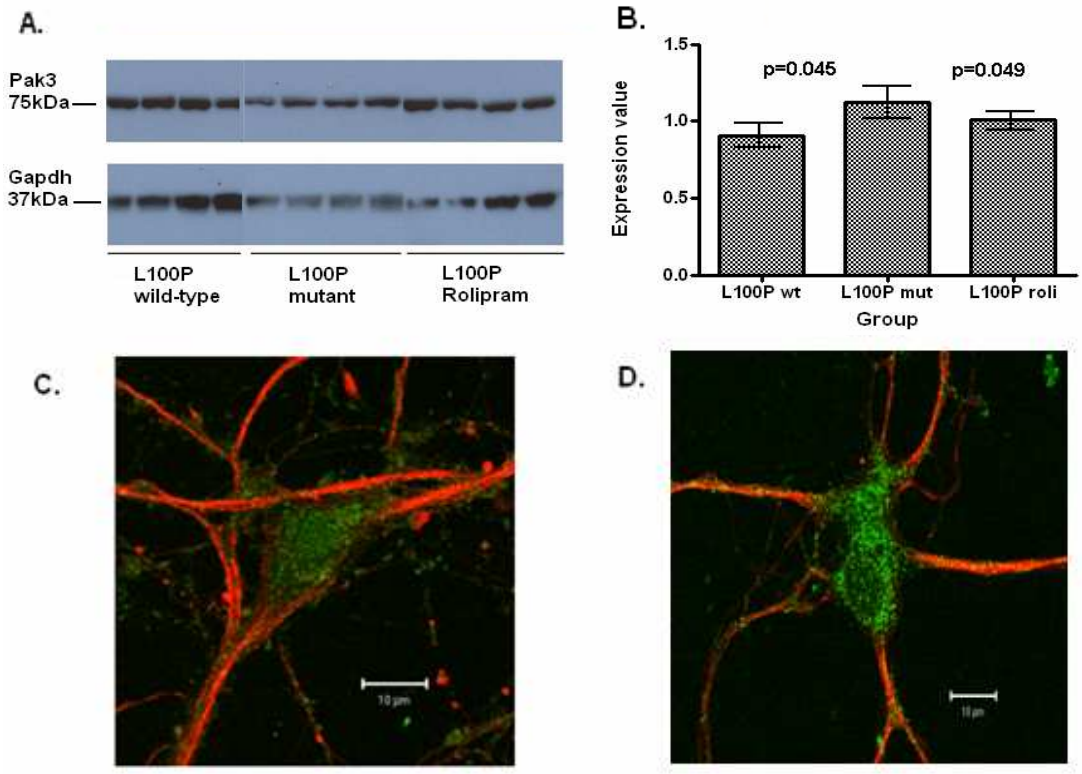

Figure 6.2: Western blotting and ICC for Pak3 in L100P ENU mutant mice. Panel A shows a representative western blot for Pak3 with Gapdh membrane used for correction below. Panel B shows the semi-quantative analysis of the blot using image J and a standard student t-test ($p<0.05$). Panel C shows the confocal microscope image of wild-type primary cultured neurons stained for Pak3 (green) and co-stained for β-actin (red). Panel D shows the mutant primary cultured neurons stained for Psk3 (green) and co-stained for β-actin (red).

There is a small but significant upregulation of expression at the protein level in the L100P drug naïve adult mouse compared to wild-type littermates that is corrected by treatment with rolipram ($p<0.05$). This correlates well with the results obtained from the gene expression analysis. No global changes in protein localisation are observed from ICC of primary cultured hippocampal neurons, but as the protocol used is not quantitative this cannot be accurately defined without further analysis. Initial observations suggest there may be slightly more puncta in the dendrites of the mutant animal, but further analysis and staining would be required to confirm this. The overall expression pattern of Pak3 in the hippocampal neuron is similar to that of Egr4. Further work is again required to determine whether this signifies co-localisation with Disc1.

7.3 Protein analysis of NRXN1

Nrxn1 has previously been identified as a schizophrenia candidate gene [83, 251-253] and was found to be upregulated in the L100P drug naïve adult mouse, in both the array experiment and subsequent qRT-PCR analysis. Protein which had been extracted from the hippocampus of animals used for batch two of the qRT-PCR was used to determine differences in expression at the protein level (figure 7.3). Nrxn1 is expressed as multiple isoforms due to alternate exon splicing. The two classes of isoform are determined by splice site and size. Large, or alpha, isoforms are the most common of the two. The most widely expressed Nrxn1 isoform is around 150kDa and the antibody used in this study was designed to primarily detect this isoform, although it would also detect other large isoforms of Nrxn1. One replicate was carried out by Rosie Walker and the rest by myself.

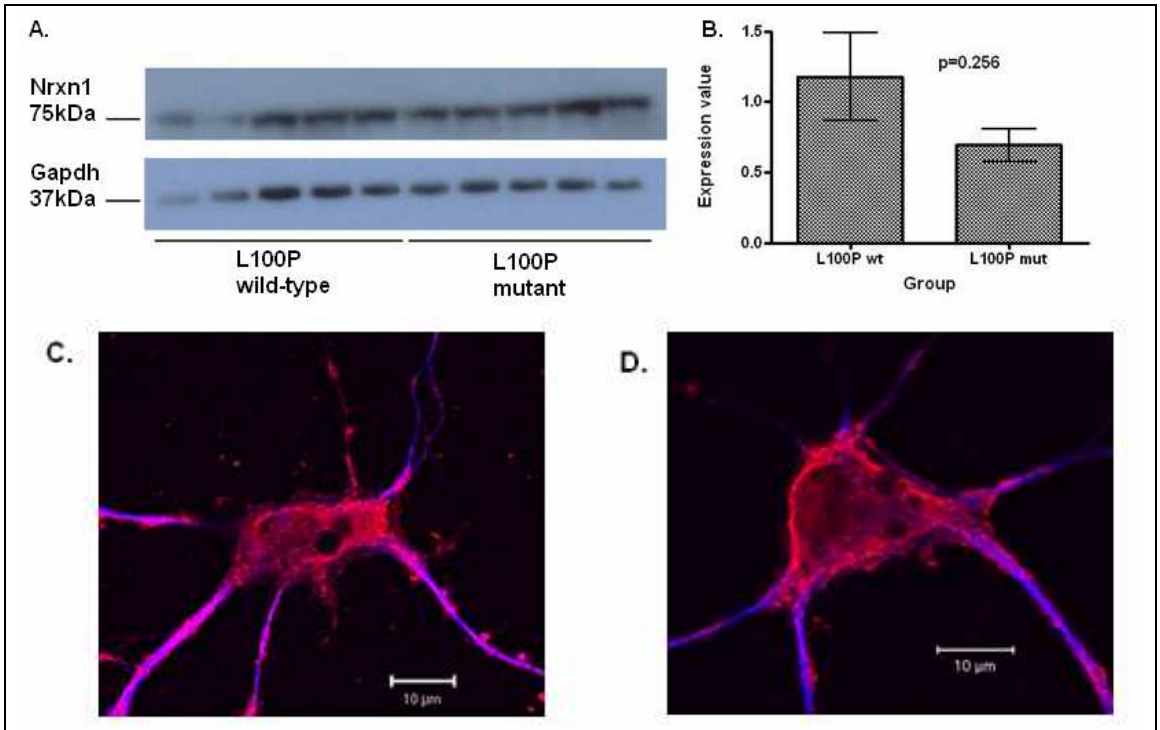

Figure 7.3: Western blotting and ICC for Nrxn1 in L100P ENU mutant mice. Panel A shows a representative western blot for Nrxn1 with Gapdh membrane used for correction below. Panel B shows the semi-quantative analysis of the blot using image J and a standard student t-test ($p>0.05$). Panel C shows the confocal microscope image of wild-type primary cultured hippocampal neurons stained for Nrxn1(red) and co-stained for β-actin (blue). Panel D shows the mutant primary cultured hippocampal neurons stained for Nrxn1(red) and co-stained for β-actin (blue).

The trend of expression of Nrxn1 protein between the L100P drug naïve adult mouse and the wild-type littermate is in the opposite direction to that observed in both the array and qRT-PCR analysis, however no significant difference was observed ($p>0.05$). Additionally, the datasheet for the antibody used stated the band size as 150kDa and not the 75kDa observed. After consulting the technical team at Abcam who supplied the antibody, they have stated the 75kDa band is a Nrxn1 isoform, but have removed the mouse from their 'tested in' species. In the future this experiment should be repeated with a better characterised antibody. This will be discussed further in section 7.6. Again no global changes in protein localisation are observed from ICC of primary cultured neurones but as the protocol used is not quantitative this cannot be accurately defined

without further analysis. The pattern of expression suggests endoplasmic reticulum involvement but further staining of cellular components would be required to substantiate this.

7.4 Protein analysis of Nrxn3

Nrxn3 is involved in synaptic transmission and neurotransmitter secretion, and has previously been associated with alcohol and nicotine dependence [255]. It was found to be upregulated in the L100P drug naïve adult mouse, in both the array experiment and subsequent qRT-PCR analysis. Protein which had been extracted from the hippocampus of animals used for batch two of the qRT-PCR was used to determine differences in expression at the protein level (figure 7.4). Like Nrxn1, Nrxn3 also has multiple isoforms due to alternate splicing. The antibody used was designed to detect alpha isoforms of around 150kDa. One replicate was performed by Susan Anderson.

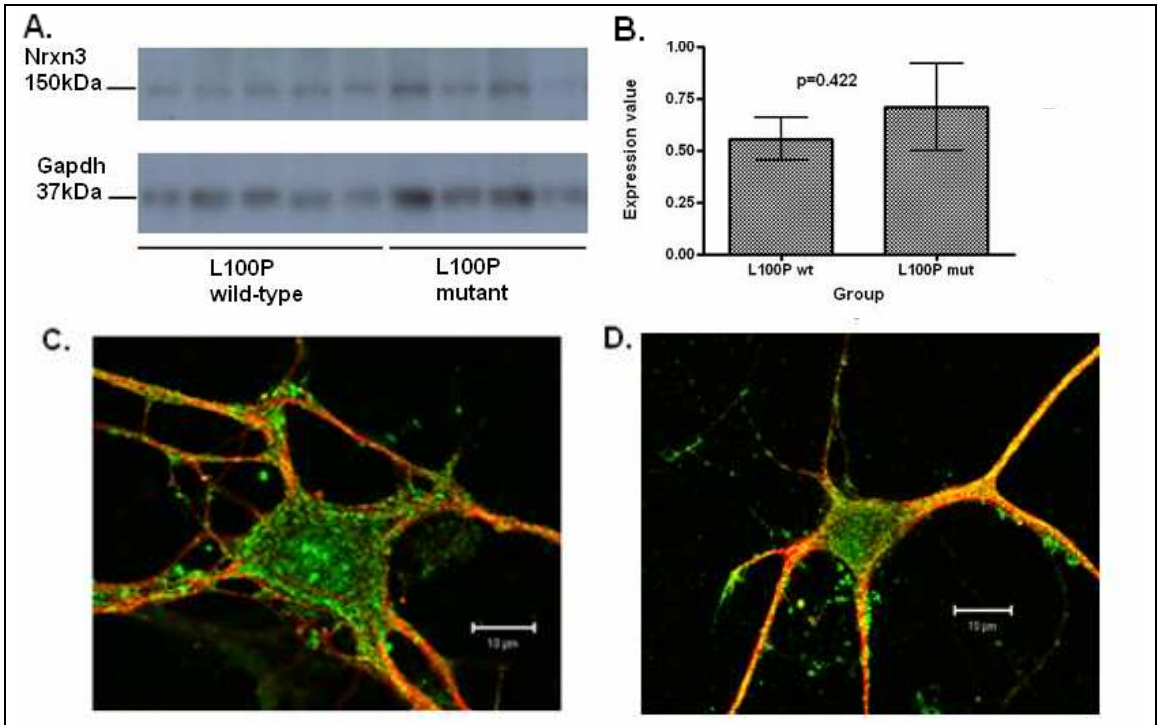

Figure 7.4: Western blotting and ICC for NRXN3 in L100P ENU mutant mice. Panel A shows a representative western blot for Nrxn3 with GAPDH membrane used for correction below. Panel B shows the semi-quantative analysis of the blot using image J and a standard student t-test ($p>0.05$). Panel C shows the confocal microscope image of wild-type primary cultured hippocampal neurons stained for Nrxn3 (green) and co-stained for β-actin (red). Panel D shows the mutant primary cultured neurons hippocampal stained for Nrxn3 (green) and co-stained for β-actin (red).

While the trend of expression is consistent with the array and qRT-PCR analysis, no significant difference was observed at the protein level between the L100P drug naïve adult mouse and the wild-type littermate ($p>0.05$). The blot was considered rather dirty and it was difficult to obtain a clear result from this antibody. Another Nrxn3 antibody was tested but the result was even more ambiguous and this was considered the better of the two for analysis, however, these results should be viewed as preliminary. Again no global changes in protein localisation are observed from ICC of primary cultured neurones but as the protocol used is not quantitative this cannot be accurately defined without further analysis.

7.5 Protein analysis of Sort1

Sort1 (neurotensin receptor 3) has previously been found to be down-regulated in sufferers of bipolar disorder (Christoforou et al, manuscript in preparation) and has been implicated in the modulation of dopamine signalling [270]. It was found to be down-regulated in the L100P drug naïve adult mouse pre outlier removal, and remained significant by qRT-PCR analysis post outlier removal. Sort 1 is expressed as a 95kDa protein (figure 7.5). One replicate was performed by Susan Anderson.

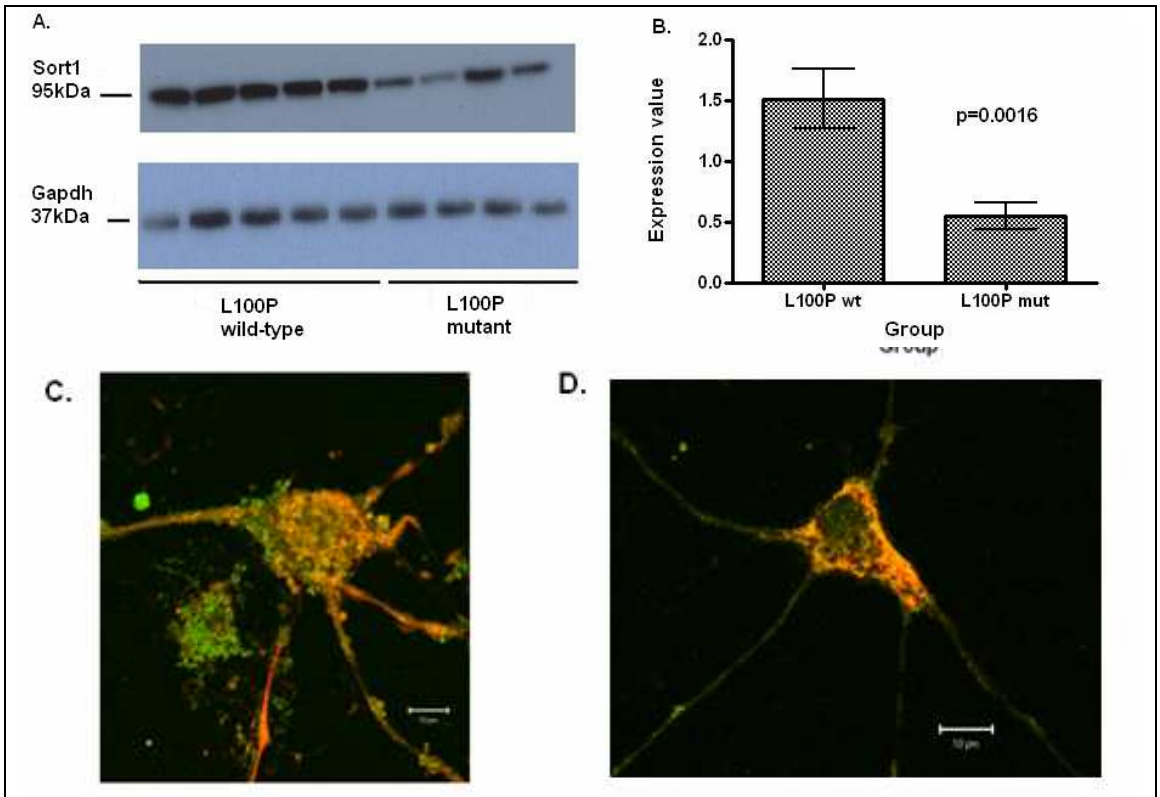

Figure 7.5: Western blotting and ICC for Sort1 in L100P ENU mutant mice. Panel A shows a representative western blot for Sort1 with gapdh membrane used for correction below. Panel B shows the semi-quantative analysis of the blot using image J and a standard student t-test ($p<0.01$). Panel C shows the confocal microscope image of wild-type primary cultured hippocampal neurons stained for Sort1(green) and co-stained for β-actin (red). Panel D shows the mutant primary cultured hippocampal neurons stained for Sort1(green) and co-stained for β-actin (red).

There is a significant difference at the protein level between the L100P drug naïve adult mouse and the wild-type littermate ($p<0.01$). Images obtained from the ICC of Sort1

appeared blurry with bleaching of the Sort1 signal and the Actin signal. This occurred in every image taken and was consistent between wild-type and mutant, suggestive of an incompatibility between the two antibodies. Again no global changes in protein localisation are observed from ICC of primary cultured neurones but as the protocol used is not quantitative this cannot be accurately defined without further analysis.

7.6 Conclusions

While identification of dysregulated genes gives insight into the mechanisms of mental illness, correlation at the protein level signifies a lack of cellular correction mechanisms allowing consequential effects of the gene disregulation at the cellular level. In a review by Chuaqui et al (2002) [277]they state that a correlation between changes at the mRNA level and the protein level are accurate less than 50% of the time. This is due to post-translational modifications of the proteins, and cellular feedback mechanisms which maintain proteins at optimal levels for cellular function and development.

This study has identified three genes that are significantly differentially expressed at both the gene and protein level; *Egr4*, *Pak3* and *Sort1*. These three genes have all been previously implicated in major mental illness. Both *Nrxn1* and *Nrxn3* failed to replicate their differential expression at the protein level. This could be due to the mechanisms mentioned previously, however there were also problems involved with the NRXN antibodies used which could account for the lack of change observed. NRXN1 is present in multiple isoforms and non-specific binding could account for the problems encountered with this antibody. As mentioned in section 7.3, the NRXN1 antibody did not give a band corresponding to that suggested on the manufacturers' datasheet. Consultation with the Abcam technical team revealed the antibody had not been tested in this species, and it could not be guaranteed that this was specific binding. A clear result could not be consistently obtained for the NRXN3 antibody. There are multiple isoforms of both NRXN1 and NRXN3 that could potentially bind the antibodies used and obtaining reliable results proved problematic.

Protein concentration was measured for all samples prior to loading using the Thermo Scientific NanoDrop 100 spectrophotometer. Despite loading what was believed to be the same amount of protein (after quantification and required dilutions) in each well, the GAPDH loading control did not give the same intensity band for each well. There was no consistency in the direction of difference (ie the mutant animals were not consistantly higher than the wild-type with regards to protein concentration) nor was the position on the gel a factor. It is likely that these inconsistencies were due to errors with the initial quantification or inaccurate pipetting when loading the gel. It would be advisable to test the calibration of the NanoDrop for protein quantification by running samples on both the NanoDrop platform and a Bradford or a Lowry assay to check for inconsistencies. Due to these loading inconsistencies there was a greater margin for error when correcting the band intensity for overall protein concentration by densitometry. The outcome of western blot densitometry relies highly on the equipment used at every stage. In 2009, Gassmann *et al* [278] discuss the pitfalls of current methods of densitometry and state that the same blots can give vastly different *p*-values (0.000013 to 0.76) depending on the method of digitization used. By comparing the densitometry results obtained from the same blot using a densitometry camera and a standard office scanner (as used in this experiment) they concluded that the office scanner was not a suitable tool for digitizing the image due to the inability to switch off the gain control. This resulted in a short optical density range and saturation at the higher levels was more frequent than with the densitometry camera. Lack of homogeneity across the scan area also means that the optical density of a band or spot on the gel is a function of its position, and shading correction that would correct this problem cannot be performed due to the gain control. In this experiment, the image J software used relied on the quality of scanner used when uploading the blot image to the PC, and on the blots being compared being of a similar exposure. Grassmann *et al*'s results would suggest that a different digitization technique would have resulted in very different results from the same blots. As such the western blots displayed here may only be taken as an indication of differential expression and not an absolute value. Further attention to quantification methods may or may not confirm the reported result and it is suggested this be carried out using other suggested methods.

A similar statement could be made for the immunocytochemistry staining reported in this chapter. There were no observable differences between mutant and wild-type animals in relation to protein distribution in the neuron, however this was a very broad brush approach aiming to identify areas for further study. Further research should involve co-staining the protein of interest with a synaptic marker, such as PSD-95, to determine effects on synapse localization. Of interest would also be co-staining with markers for other cellular components (such as mitochondria etc) to give a greater idea of the effect of the dysregulated proteins on cellular function as a whole, and their localization in relation to Disc1.

The above results, however, provide a good grounding for further research into the expression of these candidate genes at the protein level, in particular the localization and expression in vitro.

Chapter 8
Discussion

8.1 Summary of findings

This thesis has described findings from genome wide expression analysis, qRT-PCR and protein analysis of an accepted animal model of major mental illness. I have reported the discovery of a number of potential candidate genes for further research. In this final chapter I will present a summary of the findings, how they relate to the field of mental illness research, and a possible direction for future work based on these findings.

8.1.1 Chapter 3: Sample collection and processing of a genome wide microarray analysis of the Disc1 ENU mouse mutants

One of the primary driving forces behind this study was the identification of the L100P and Q31L mouse mutants, and subsequent behavioural data classifying them as 'schizophrenia-like' and 'depressive-like' [197]. Both of these models carry missense mutations in exon 2 of the *Disc1* gene which disrupt Pde4 binding sites. Expression of Disc1 protein is not altered in either animal, however Disc1-Pde4b binding is reduced in both models, and activity of Pde4b is reduced by up to 50% in the Q31L model [197].

I repeated the prepulse inhibition studies carried out by Clapcote *et al* (2007) [197] in an attempt to confirm the behavioural phenotype in the mouse model prior to genome wide expression analysis. The prepulse inhibition experiments I carried out did not give statistically significant results, but did follow the trend previously published. However, as mentioned in section 3.8, there were a number of caveats associated with these experiments. The identification of a *de novo* mutation in the mutant mouse lines meant it was necessary to cross only homozygous animals, making the use of wild-type littermates as controls impossible. As the mutation was on a C57BL6 background, this strain was used as a wild-type control. Studies of PPI in rodents have been going on for many years as researchers attempt to find the genetic basis for such behaviours. In 1997, Paylor and Crawley studied the differences in PPI in commercially available inbred mouse strains. Interestingly, the C57BL/6J mice gave the lowest PPI of all the mice they tested[232]. The PPI of the C57BL6 mice used in this study was not statistically different to the

results obtained by Paylor and Crawley, but was significantly lower than Clapcote *et al*'s wild-type controls. The result obtained for the mutant mouse lines are not significantly different to those obtained by Clapcote *et al* (2007) in the mutant lines, suggesting that, at least for the behavioural study, the C57BL6 mice are a less than ideal control animal. It was concluded that, due to the similarity between the levels of PPI I observed and those obtained by Clapcote et al (2007) [197], that the PPI deficit in the mutant lines was a true result and the behavioural phenotype was present in the mice. In addition to this, other researchers have confirmed the PPI deficit in the ENU mouse lines using wild-type controls (Steven Clapcote, personal communication).

Despite the limitations, the genome wide gene expression analysis was carried with C57BL6 control animals out due to time pressure and because validation of differential expression would be carried out using wild-type littermate controls once a colony had been established. The microarray analysis, therefore, compared drug naive adult mutant mice or adult mutant mice treated with selected antipsychotic or antidepressant drugs, with C57BL6 or drug naïve controls. Samples were also taken from embryonic day 13, which had previously been shown to be a peak in Disc1 protein expression [134]. Quality of RNA extraction from hippocampal tissue (or whole brain from the embryonic stage) was tested using Agilent technology. This showed that the RNA extracted was of sufficient quality and could therefore be used to look for differential gene expression.

8.1.2 Chapter 4: Differential Expression Analysis of the Disc1 ENU Mutant Mouse Microarray

Analysis of the genome wide expression data initially revealed a total of 835 genes that were differentially expressed across the sample set. However, subsequent analysis of individual gene expression using statistical methods and qRT-PCR identified one outlier pool that gave erratic and unpredictable expression values. Re-analysis of the whole genome data without this pool as expected revealed a reduced set, with 368 genes differentially expressed across the sample set. Of these genes, 62 were over-expressed

and 24 under-expressed in the L100P adult samples and 19 were found in the Q31L adult sample set. Three genes (*wdfy1, LOC674214* and *2900040C04RIK*) were found to be differentially expressed in both the L100P and Q31L non-drug treated adult mice. 91 genes were over-expressed and 95 genes under-expressed in the Q31L bupropion treated group compared to non-drug treated adults. The L100P rolipram treated mouse gave the most genes which showed correction under pharmacological influence. Of the 32 genes identified, seven showed correction of the L100P non-drug treated disregulation. Correction of expression was defined as returning that gene to a level not significantly different (FC+/- 1.3, $p<0.05$) to that of C57BL/6J controls. 45 genes were found to be differentially expressed in the L100P embryo and 25 genes differentially expressed in the Q31L embryo. Four genes were identified in both groups, *Prm3, Coro1B, Numb* and *Pde4d*. With the exception of bupropion treatment, the L100P comparisons all produced a higher yield of dysregulated genes than the Q31L comparisons. The 'schizophrenia-like' phenotype of the L100P mouse is considered the more severe of the two, and so it could be expected that this would give a higher level of differential expression observed in this animal. Due to the high number of genes differentially expressed with bupropion treatment that were not correspondingly differentially expressed in the Q31L mutant without treatment, it was concluded that this was an effect of drug treatment and not related to the genotype. No differences were observed between male and female mice in gene expression. The small number of genes that were differentially expressed in the adult mutants and then corrected by drug treatment would suggest the drugs used were active primarily at the protein function level, and had minimal effect on gene expression.

GOTree analysis revealed over-enrichment of genes involved in the following categories: cell-cell signaling, transmission of nerve impulses/synaptic transmission. No overenrichment of gene categories was observed in any other group tested, possibly due to the small number of differentially expressed genes. Previous studies have identified clusters of genes involved in neural function [199] synaptogenesis and sensory perception [128] in schizophrenia. From the GOTree analysis, *Egr2* and *Egr3*, which have previously been associated with schizophrenia in a Japanese population, [238] were both significantly down-regulated in the L100P adult sample, and are involved in cell

signaling and transmission of nerve impulses. The excitatory glutamate receptor 2 (*Gria2*) was overexpressed in the L100P adult and is involved in cell signaling, transmission of nerve impulse and neurotransmitter secretion. It has also been identified as a potential candidate gene for schizophrenia in its own right [279]. Synapsin 2 was overexpressed in the L100P adult and was identified as being involved in cell signaling, neurotransmitter secretion, and transmission of nerve impulses. *Syn2* has previously been implicated in confering risk to schizophrenia in a Korean cohort [280]. Neurexin 1 (*Nrxn1*) was overexpressed in the L100P adult mouse. It is involved in maintenance of synaptic junctions and is a mediator of intracellular signalling[249]. *NRXN1* was implicated by Kirov *et al* (2008) [217] in an analysis of copy number variants in schizophrenia. Other studies on copy number variants have also implicated *NRXN1* in schizophrenia pathology. Need *et al* (2009) [251] found deletions in the 3' end of *NRXN1* in patients with schizophrenia, but not in control groups. Vrijenhoek *et al* (2008) [252] suggest that CNVs that affect the first few exons of *NRXN1* confer greater risk of major mental illness, while Rujescu *et al* (2009) [253] suggest that deletions that affect exons directly increase susceptibility. In addition, a missense mutation in exon 1 of human *NRXN1* has also been linked to autism [254] and recent studies have linked deletions in *NRXN1* to a range of phenotypes, including autism spectrum disorder, mental retardation, hypotonia and language delays [281].

Lots of recent work has investigated the role of copy number variants (CNVs) in neurodevelopmental disorders (reviewed in Kirov 2010) [82]. Only two genes were present in both the dysregulated gene list from this study and copy number variants in the previous studies. *MLL5*, a transcription regulator, was present as a duplication in schizophrenia cases from Walsh *et al* 2008 [83], and was upregulated in the L100P non-drug treated adult group. As previously mentioned, *NRXN1* was also found to be up-regulated in the L100P non-drug treated adult group. Walsh *et al* (2008) reported a *NRXN1* deletion present in a schizophrenia case and affected sibling [83]. The deletion spans the promoter and first exon of the gene and partially overlaps with previous deletions co-segregating with autism [76] and mental retardation [239]. I also had access to data from recently completed microarray expression studies on LCLs from individuals

from the *DISC1* translocation (t1:11) family that allowed an almost direct comparison to be made between the mouse and human model. Five genes (*Dusp6, F5, Hook3, Nrxn1 and Nrxn3*) were differentially expressed in both studies, with *Hook3, Nrxn1* and *Nrxn3* dysregulated in the same direction in both arrays. These overlapping results between the mouse and human sample subjects is promising in terms of candidate gene identification and because it provides further evidence of the validity of the animal model to this system.

One major flaw of this comparison is the two arrays used different tissues as the primary RNA source. While it was possible to use brain tissues from the ENU mutant mouse, the human samples were from a blood-derived cell line. Blood-derived (lymphoblastoid) cell lines do not express all genes and in particular those expressed in the brain[240]. This could obviously result in some genes not being present in the human dataset by virtue of the original tissues used, and not because they are not expressed, or potentially dysregulated in the human condition.

Using Ingenuity pathway analysis, nine genes were identified that showed both overenrichment in GO categories and presence in the top statistically significant Ingenuity pathway analysis network in the L100P adult group. Ingenuity pathway analysis identified networks involved in cellular development, gene expression and cell signaling for the L100P adult, L100P embryo and Q31L embryo respectively. On the basis of fold-change, p-value, function (by GeneOntology classification) and overlap with previous studies, 40 genes from the microarray study were chosen for validation by qRT-PCR. Genes chosen were mainly involved in synaptic transmission, neurotransmitter transport and secretion, and central nervous system development. Genes whose differential expression was corrected by drug treatment were also carried forward for validation.

This study had obvious limitations and weaknesses, not least of all the presence of an outlier group in the initial analysis. In addition, a number of the genes that showed differential expression on the array were anonymous sequences from a cDNA bank with

little or no functional information attached. This made it difficult to determine the relevance of these predicted sequences to major mental illness and therefore, most were not chosen for further analysis. It is possible that the lack of information available for many genes biased the selection of the subset for further analysis.

8.1.3 Chapter 5: Quantitative real-time PCR analysis of candidate differentially expressed genes in the L100P and Q31L mutant mice

qRT-PCR was used for the validation of the microarray results. Two cDNA batches were used; one from the RNA that went on the array, and an independent batch extracted from mutant animals and wild-type littermate controls. Genes that failed to validate from the first batch of cDNA were not tested in the second batch. While every effort was made to ensure overlap of the microarray probes with the qRT-PCR probes, a perfect match could not be ascertained. In all cases both probes were derived from the same exon but, as the sequences of the qRT-PCR probes are not made publicly available, identical sequence alignment could not be achieved. Slight misalignment of probes could result in amplification of different isoforms of the gene in question. Also, false positivity in array data can occur by binding of repetitive nucleotide elements, sequence homology between functionally different transcripts and high background levels due to nonspecific binding of nucleotides [248]. For these reasons it was decided that genes must show consistent results across platforms to be considered valid, thus ensuring as much as possible a true result.

Of the 40 genes tested by qRT-PCR, 24 validated in the first round and only seven gave robust and reliable results through both rounds of validation. Correlation estimates predict a Spearman coefficient of 0.708 [282] between microarray data and qRT-PCR. The results obtained from this study are therefore slightly lower than would be expected, however lack of concordance between platforms at low fold-change (<1.4) has been previously reported [283]. What is more concerning is the lack of concordance between the two batches of RNA, suggesting differences in the wild-type expression could be due to minor genetic differences between the C57BL/6J and the L100P wild-type. Although

the L100P animals had been backcrossed for over 10 generations it is unlikely, but not impossible, that there were still some genetic differences present.

Of the seven genes validated, four were from the L100P drug naïve adult group (*Nrxn1, Nrxn3, Egr4 and Wdfy1*), two from the L100P embryonic group (*Cdh11, Prm3*) and one from the L100P Rolipram treated group which showed correction (*Pak3*). The neurexin genes (*Nrxn1* and *Nrxn3*) are involved in synaptic transmissions and the maintenance of synaptic junctions. *Nrxn1* has been shown to be involved in the maturation and differentiation of GABAergic and glutamatergic synapses through bi-directional signalling [250] . DISC1 has previously been shown to interact with GSK3β, which suppress long-term potentiation and presynaptic release of excitatory glutamate in cortical neurons by inhibition [284]. As the glutamate hypothesis of schizophrenia predicts that decrease of NMDA receptor signaling during interneuron development could result in the behavioural phenotypes observed in schizophrenia [285], the potential interaction between *DISC1* and another regulatory glutamate gene is of great interest. *NRXN1* has also previously been identified as a candidate gene for schizophrenia [83, 217, 251-253] and Autism [254] while SNPs in *NRXN3* have been associated with alcohol and nicotine dependence and linked with opiate dependence [255, 256]. The Early Growth Response Factors (*EGR*s) 2, 3 and 4 are synaptic activity inducible immediate early genes and all show nominal association with schizophrenia in a Japanese population. *EGR1, 2* and *3* are downregulated in the prefrontal cortex of schizophrenic brains [238] , *EGR4* expression was too low to measure disregulation however, treatment with the atypical antipsychotic and antidepressant Aripiprazole increases expression of *EGR4* in rat frontal cortex[257]. Little is known about the function of WD repeat and FYVE domain-containing protein 1 (WDFY1 or FENS1) but it is known to be ribosomal and has been associated with alcohol consumption and preference [258].

Cadherin 11 (*Cdh11*) is a type 2 classical cadherin involved in mediation of calcium dependant cell adhesion. It was first discovered in rodent brain samples [266] with high expression in the developing olfactory system at E13-E17 [267] and has since been linked to multiple cancers [268, 269].

Pak3 was the only drug-corrected gene validated. Mutations in *PAK3* have been associated with X-linked nonsyndromic mental retardation [260-262] and suppression of *Pak3* results in formation of abnormally elongated dendritic spines and a reduction of mature synapses [263]. *Pak3* knockout mice have deficiencies in learning and memory and abnormalities in synaptic plasticity along with a reduction of transcription factor cAMP-responsive element-binding protein (CREB) suggesting a novel signalling mechanism with PAK3 and Rho signalling regulating synaptic function and cognition [264]. *PAK* genes also interact with the known DISC1 interactor Kalirin-7 which is thought to control multiple aspects of spinal plasticity [265]. Kal-7 activates Rac-1 to control spine size. In the presence of DISC1, Kal-7 is anchored, regulating access to Rac-1 and controlling duration and intensity of Rac-1 activation. It is thought that constitutive activation of Rac-1 (due to Kal-7 not being anchored by DISC1) results in decreased spine size and may underlie the disturbances in glutamatergic neurotransmission observed in schizophrenia [286].

Five genes from the pre-outlier removal gene list (as described in chapter 5) were tested by qRT-PCR and two validated through both rounds, with significant fold-change and p-values that meet the criteria imposed. As previously mentioned, false positivity in array data can occur through multiple routes. In parallel, false negativity can occur through low expression levels, inefficient priming of specific mRNAs resulting in transcript drop out, poor adhesion of DNA to the slide, and splice variants with sequences not included on the array [248]. As the fold changes observed on the microarray for these genes were below the accepted sensitivity level of the array (http://www.switchtoi.com/pdf/GXHuman6-8v2Datasheet.pdf) but were subsequently found to be differentially expressed through qRT-PCR they may be considered as false negatives. This does, however, raise speculation to the criteria imposed when selecting genes for validation and this will be discussed in section 8.2.

Three additional genes (*Dusp1, Ndfip* and *Atp5b*) were found to be robustly dysregulated in the L100P rolipram treated group, but did not differ in the L100P adult vs wild-type controls. I, therefore, proposed that the disregulation of these genes was a function of drug treatment and not a result of the mutation. To confirm this I tested these genes in a

C57BL6 drug naïve group vs C57BL6 rolipram treated group. *Dusp1* was significantly up regulated in the C57BL6 rolipram treated group compared to the drug naïve group, as it had been in the L100P rolipram treated group compared to the L100P drug naïve group. *Ndfip* also behaved in the same way in the C57BL6 rolipram treated group as it had in the L100P drug treated group, being significantly down regulated compared to the drug naïve control. *Atp5b* was significantly down regulated in the C57BL6 rolipram treated group where it had been up regulated in the L100P rolipram treated group, but a significant difference in expression was observed suggesting it is an effect of the drug treatment as opposed to the mutation. As the changes in expression in these three genes appear to be an effect of the rolipram treatment rather than the mutation I feel they do not fit the criteria for further analysis within this project. These data may, however, provide some information on the pathways through which these drugs act.

8.1.4 Chapter 6: Establishing a developmental profile for the genes of interest in L100P mutant mice and wild type controls

The developmental profile of *Disc1* highlighted key stages of development where expression is altered in the mutant mouse line, and also confirmed my previous microarray result which showed that *Disc1* expression was not significantly different at either E13 or adulthood between the L100P mutant and wild type controls. It was previously thought that the ENU mutations affected DISC1 protein function and not expression levels [197]. The data from these developmental profiles redefines our knowledge of this model and suggests that the behavioural phenotype observed is a combined effect of differences in DISC1 function and developmental expression.

In the L100P mutant mouse there is an obvious peak in *Disc1* expression at P1, with a peak at E18 observed in the wild-type controls. This is similar to the pattern observed in the Q31L mutant mouse. In addition, the Q31L mutant animals have significantly higher *Disc1* expression than their wild-type counterparts at E15. It can be speculated that the

different developmental expression patterns of the Q31L and L100P mutants may contribute to the different phenotypes observed in behavioural testing. A review of the literature on mouse developmental stages suggests that, based on genes with known neural functions, the functional gene clusters most highly expressed at E15 are involved in neural determination and differentiation, and at E18 and P1 are involved in synapse formation and function, and survival and growth [274]. Based on all genes differentially expressed at E18 and P1 the top functions are intra/inter cellular molecular transport, and signal transduction [274]. This delay in *Disc1* expression in the L100P mutant mouse could have detrimental effects on neurodevelopment in the mutant animals. Personal observations and those of other researchers working with these animals show that the mutant individuals are physically smaller when born, which could signify a delayed development (Steven Clapcote personal communication). Activity at the three developmental time points highlighted here roughly correspond to human neurodevelopment at weeks 9-12, 12-15.5 and 13.5-17.5 *in utero* [287].

While the result of one gene expression profile is not enough to stipulate cause and effect it is worth noting that low birth weights have previously been associated with schizophrenia [273] and a latent shift in the expression of key developmental proteins could be a factor in this. Hikida *et al* [193] also note that there is a 3month delay in neuronal migration in a *Disc1* mouse model, which could potentially be attributed to a change in *Disc1* expression during development.

Developmental expression analysis of five differentially expressed genes highlighted some interesting developmental changes. Both neurexin genes showed a marked shift of peak expression in the L100P mutant compared to wild-type littermates. *Nrxn1* peaks at E18 in wild-type, and P1 in L100P mutant animals. Conversely, *Nrxn3* plateaus from E15-P1 in wild-tpe and peaks at E15 in mutant animals, with significantly lower expression at E18 and P1 than wild-type. As the predicted developmental function at E18 is the formation of synapses, the dip in expression of genes involved in synaptic maintenance and transmission in the mutant animals is considered of key importance. *Pak3* developmental expression is identical in the wild-type and mutant animal until E15.

After this the wild-type expression continues to rise to peak at E18, while the mutant expression decreases, plateauing at P20. There is significant underexpression of *Pak3* at E18 and P1 in the mutant mouse. As *Pak3* was the only gene to show correction through drug treatment, it would be interesting to see if earlier intervention would facilitate complete rescue of the phenotype, or if treatment post hoc is sufficient. *Egr4* displayed a distinct lack of expression in both wild-type and mutant until P1, which has previously not been reported, and significantly higher expression in the L100P mutant at P20, just prior to puberty in the mouse. According to the developmental hypothesis of schizophrenia, the molecular changes in the brain at the time of puberty are partially responsible for disease development in predisposed individuals[189]. If this is the case then the altered expression of this and other genes studied during development could be responsible for deficits in key circuitry which are exposed at this stage, facilitating the disease phenotype, as described in Weinberger's developmental hypothesis [189], in the L100P mutant model.

8.1.5 Chapter 7: Preliminary protein analysis of validated genes in L100P drug naïve (saline) adult mouse samples

Due to post-translational modifications of proteins, and cellular feedback mechanisms that maintain proteins at optimal levels for cellular function and development, it has been reported that a correlation between changes at the mRNA level and the protein level are accurate less than 50% of the time [277]. I identified three genes that are significantly differentially expressed at both the gene and protein level; Egr4, Pak3 and Sort1 in the L100P adult mouse model. Both Nrxn1 and Nrxn3 failed to replicate their differential expression at the protein level. This could be due post translational modifications of the proteins, however there were also problems involved with the neurexin antibodies used which could account for the lack of change observed. The Nrxn1 antibody did not give a band corresponding to that suggested on the manufacturers' datasheet and clear result could not be consistently obtained for the Nrxn3 antibody. There are multiple isoforms of both Nrxn1 and Nrxn3 that could potentially bind the antibodies used and obtaining reliable results proved problematic.

Immunocytochemistry was used to visualise the proteins of interest in cultured hippocampal neurons from L100P mutant and wild-type animals. No obvious overall changes in expression pattern were observed in any of the cultures, however the analysis was not quantitative and small changes would not be observed by the naked eye. Egr4 appears to be expressed in disperse puncta on the dendrites, with the majority of Egr4 expression contained within the soma. Disc1 has been shown to be expressed in similar puncta in the dendritic shaft with localisation to the ribosomes and microtubule structures in the soma [276]. It can be speculated that Egr4 has a similar, although not identical, expression pattern to Disc1 in the neuron, although further staining of specific cell components would be required to confirm this. The overall expression pattern of Pak3 in the neuron appears similar to that of Egr4. Initial observations suggest there may be slightly more puncta in the dendrites of the mutant animals but further quantitative analysis would be required to confirm this. Nrxn3 again appears localised to the some, with few puncta in the dendrites. Staining of the hippocampal neurons for Nrxn1 gave a slightly different pattern of expression, suggestive of endoplasmic reticulum involvement. Disruption of the neurexin PDZ-binding domain motif in Nrxn1 knockout mice has been shown to result in diffuse distribution patterns with Nrxn retained in the endoplasmic reticulum and not packaged into vesicles for transport to the synapse [288]. It could be speculated that disruption of Nrxn1/PDZ binding by Disc1, or one of its interactors, prevents trafficking to the synapse. Further staining with synaptic markers would be required to determine if Nrxn1 is localising to the synapse or if it is being held in the ER. Images obtained from the ICC of Sort1 appeared blurry with bleaching of the Sort1 signal and the actin signal. This occurred in every image taken and was consistent between wild-type and mutant, suggestive of an incompatibility between the two antibodies.

While there were a number of limitations to this study (as described in more detail in section 7.6), these results provide the basis for further research in this area. They are to be considered preliminary findings, and it would of course be advantageous to continue these experiments, and in some cases repeat with more reliable antibodies. Further neuron

staining using cellular markers to compare co-localisation of proteins between mutant and wild-type animals would also be recommended.

8.1.6 Conclusions

This work has identified a number of genes dysregulated by missense mutations in *Disc1* that are proposed to be involved in the schizophrenia endophenotype displayed by the mouse model. Most notably, the striking disregulation of a number of known candidate genes for major mental illness in the L100P *Disc1* mutant animals suggests a common network involved in the pathogenesis of schizophrenia and major mental illness. The altered expression of these genes during development in the mouse model also suggests a substantial neurodevelopmental component, supporting previous hypotheses from Weinberger (1986) [189] and Torkamani (2010) [199].

8.2 Caveats

No experiment is perfect, and it is important to record and analyse the potential caveats to improve future work. In the whole genome expression study, potentially the biggest question was the use of C57BL6 as a wild type control, particularly given its reduced PPI compared to the mutant wild-type littermates. The justification for this was two fold; firstly, the results from the array would be validated using wild-type littermate controls so any change attributed to the use of C57BL6 over wild-types would be identified at a later date. Secondly, the mutant mouse line had been backcrossed for 10 generations and would be expected to be genetically identical to at least 98.9%. C57BL6 mice were specifically sourced from the Charles River substrain to be as close as possible to the original backcross strain used when establishing the ENU mutant colonies, however, results obtained from both the behavioural study and comparison between the C57BL6

and wild-type littermate controls would suggest the presence of non-target mutations which may be responsible for the differences observed. These could either be secondary ENU mutations that co-segregated with the *Disc1* mutation, or sections of DBA genome that remained, despite backcrossing, around the *Disc1* gene. Comparative genotyping of the whole chromosome would allow this to be determined. In retrospect, the use of wild-type littermates for the array, as was the initial plan, would have reduced the number of false positives and false negatives, and provided a 'cleaner' result from the offset. However, due to circumstances outwith my control this was not possible, and it was necessary to continue with the closest viable control that was available at the time, the C57BL6 mouse. I would suggest that further experiments of this kind use wild-type littermate controls as standard, or that the C57BL6 be fully genotyped and compared to the mutant mouse line to fully establish and genetic differences that may be present.

Hippocampal tissue was used in the array study for a number of reasons. Abnormalities in hippocampal structure, activation, organization of neurons and synapse function have been well documented in human schizophrenia patients and other mouse models [207, 214, 233, 234]. Added to the working memory deficits often associated with schizophrenia [208], it is clear there is substantial hippocampal involvement in the disorder. The use of a single brain region also removed any confounding effects of compensation in other brain regions, which may have diluted the result. As such, this study can make direct inferences as to the gene expression in the hippocampus of the adult mutant mouse. Due to researcher inexperience, it was not possible to remove only hippocampal tissue from the embryonic mouse samples, and in these cases whole brain was used. This was not ideal and it would be suggested that analysis of distinct brain regions should be carried out on the embryonic mutant mouse, particularly with reference to the developmental panel. It would be very interesting to discover if the changes of gene expression over development observed in this study are confined to one brain region, or if they are a global effect.

Microarray technology has a number of inherent limitations that cannot be ignored. Genes with low expression levels, or genes with a small change in expression level, can often be overlooked due to the background noise of a heterogenous sample. It is highly

probable that this could occur in the sample set used for this experiment. The genetics of psychiatric disorders are known to be very complex, with multiple genes likely to be involved. Illumina state confidence in detecting a gene expression fold changes greater than 1.3, essentially meaning that any change in expression less than 30% from the norm cannot be reliably detected. Prior to filtering the data based on fold-change, 78% of genes with a significant *p*-value gave a fold-change between 1.01 and 1.25. These were obviously discarded from further analysis due to the detection cut-offs imposed by Illumina. This cut-off should however be considered arbitrary. Prior to the identification of the outlier 20 genes had been tested by qRT-PCR that did not make the 1.3 fold-change cut-off on reanalysis. Of these, two validated independently by qRT-PCR, suggestive that in these cases there was differential expression that was not apparent from the array analysis. The problem of tissue heterogeneity was in part addressed by using inbred mouse strains, however even within a small region of the brain (in this case hippocampus) there are multiple cell types with different expression patterns adding to the background noise of the array. This could be overcome using microdissection of specific cell types and it would be suggested this is carried out for future experiments.

The identification of an outlier group post analysis was indeed a blow to the study. Not least it meant the total reanalysis of the dataset and questioned the validity of previous quality control measures. As the outlier did not consistently show over or under expression, its effects were diluted when analysing the data as a whole. While this may be an isolated case, it highlights how important it is to analyse each group individually, by checking a random set of genes, as well as the group as a whole.

It was stated in chapter 4 that no corrections were made for multiple testing in this dataset. False discover rates (FDRs) were calculated for each pairwise comparison. As the maximum FDR was ~30% it could be predicted that 110 of the 368 genes found to be differentially expressed would be false positives. While this value is high it was decided that continuing without correction was viable as the data would be validated on an independent platform which should see any false positives from the array removed at that stage. If this experiment were to be repeated I would suggest either analyzing the dataset using ANOVA and posthoc tests, or, if carrying out multiple t-tests as was done here, that

some multiple testing corrections be carried out. This could potentially increase the number of false negatives from the data but would produce a dataset that was more likely to be reproducible on the independent platform.

Validation of the microarray result was two fold. In the first instance, genes of interest were analysed by Taqman real-time PCR using RNA from the animals that were used on the array. Of the subset of genes analysed, 51.5% gave reproducible results across both platforms. While every effort was made to ensure overlap of the two probes, a perfect match could not be ascertained. In all cases both probes were mapped to the same exon but as the sequences of the qRT-PCR probes are not made available identical sequence alignment could not be achieved. This may account for the low levels of reproducibility. Also, as mentioned previously, the array false discovery rate was high, so it could be expected that some of the genes identified and subsequently analysed on the Taqman platform were false positives. Wang *et al* (2006) [289]compared microarray platform and Taqman real-time PCR performance and found correlation ranges between 0.45 and 0.79 between platforms. They also noted that correlations were higher in genes with greater expression levels and larger fold changes. As the majority of genes selected for validation in this study had low fold-changes (classed as <2) it appears consistant with a smaller correlation between platforms. The second round of validation used RNA from an independent set of animals derived from the original line, with wild-type littermate controls. Only those genes that had reproduced successfully in the first round were tested in this sample set. Seven of the 19 genes tested reproduced in this second round. This could not be due to between platform differences, or probe alignment issues. It is therefore assumed that the issues discussed previously regarding the use of the C57BL6 as a control affected the result. Genotyping both 'wild-type' controls would be advised if this experiment was to be repeated, although the use of wild-type littermates from the offset would be likely to reduce this problem.

As stated previously, the selection criteria for genes for further study was, with retrospect, less than ideal. Small fold changes close to the detection sensitivity of the array could have resulted in a high number of false negatives, as demonstrated by the

validation of two genes by qRT-PCR that did not meet the selection criteria on the array. If the data were to be reanalyzed it would perhaps be prudent to select more on the basis of functional significance and *p*-value with less weight being left to fold-change values. While it is accepted that there must be a cut-off level, it is suggested that due to the complex nature of schizophrenia and other major mental illness, and the predicted multiple genes of small effect, that current array technology may not yet be sensitive enough to pick up the small changes in gene expression that may contribute to disease pathology.

The inclusion of the drug treated groups, while of interest, also increased the spread of the data and the propensity for error. Increasing the number of groups increases the multiple testing burden and reduces the statistical power. It is also likely that the primary effects of the drug treatments are not at the level of gene expression, but in post-translational protein modification and/or action on neurotransmitter systems [290, 291]. Previous studies have identified gene expression changes after clozapine treatment ([292, 293] and others), however these studies used chronic treatments and whole brain tissues, increasing the heterogeneity of the sample. I would conclude that while the drug treatments were interesting they were not necessarily an essential part of this study. Further work to investigate the actions of psychiatric drugs would be of interest, and the use of this mouse model an advantage for this, but it would be of more interest to investigate at the level of the synapse for change in neurotransmitter binding and secretion, than at the gene expression level.

8.3 Relevance of this study to the field

Since the discovery of the *DISC1* gene in 2000 [56], a number of studies have identified variants within this locus as susceptibility factors for major mental illness. Many DISC1 binding partners have also been associated with major mental illness in their own right

[131, 144-146, 156, 294], suggesting a network of gene involvement in these synaptopathies.

While microarray analysis has been previously used to identify candidate genes for major mental illness, this is the first study to look at genome wide gene expression in a model system with mutations in *Disc1*, an already established candidate gene. The use of a well classified animal model, with mutations in *Disc1* and behavioural characteristics consistent with a schizophrenic or depressive-like endophenotype, has allowed me to carry out this analysis on a uniform genetic background, with sufficient numbers for statistically relevant conclusions. Due to problems with tissue access and genetic heterogeneity, this would not have been possible using human samples.

I have identified a number of genes that are dysregulated as a result of the *Disc1* mutation. Most of these genes have not previously been shown to interact with *Disc1*, suggesting either indirect interactions and/or the possibility of network involvement of these genes, which together contribute to the disease pathology. It may also suggest involvement of transcriptional control. As some of the genes identified and validated in this study have previously been associated with major mental illness in their own right, there is a convincing argument for network involvement in schizophrenia pathology, at least in this system.

It has previously been reported that genes related to CNS development, neuron guidance and neurotransmitter secretion were down-regulated with age (birth-20 years) in normal individuals [200]. Torkamani *et al* (2010) [199] found that while this downregulation continued in control subjects, it did not in individuals with schizophrenia, suggesting a progressive neurodevelopmental deficit. This is not a new idea, as Weinberger (1986) [189] proposed the neurodevelopmental hypothesis of schizophrenia on the basis that altered gene expression during development could be responsible for deficits in key circuitry, which are exposed when molecular changes occur in the brain during puberty, facilitating the disease phenotype. As schizophrenia is not normally diagnosed until the late teens at the earliest, gene expression measurements of individuals with schizophrenia during their early post-natal development is not possible, however, using the L100P

Disc1 mutant I have shown that there are developmental gene expression differences between the L100P mutant and wild-type littermates around the time of birth, and at P20. This is an exciting development as it not only redefines what we thought we knew about the mouse model, but can be related to an already established hypothesis in a way which may help to broaden our understanding of disease development. While only a small number of genes were tested in the L100P developmental profiles, this evidence suggests a solid neurodevelopmental basis for the schizophrenia-like phenotype exists in this model system.

8.4 Differential networking and other future work

As schizophrenia and other major mental illnesses are considered complex trait disorders, the identification of individual candidate genes only gives us a small insight into the disease cause. A more general, and perhaps more useful, tool is the identification of networks of genes involved in the disease pathology. I presented data from Ingenuity Pathway Analysis, identifying predicted gene networks on the basis of fold-change and p-value from the differential expression analysis of the microarray. Networks identified were cellular development, gene expression and cell signaling in the L100P adult, L100P embryo and Q31L embryo respectively. These proposed networks rely on the expression differences between affected and unaffected individuals (or in this case, mutant and wild-type animals). Disregulation of cell-cell signaling has previously been implicated in schizophrenia [295] and Disc1 has been shown to be involved in cellular development and neurite outgrowth ([160, 296] and others) so these networks are comparable with that data. What is more interesting, is the failure of all but two of the genes contained in these networks to validate by qRT-PCR in the mouse model, suggesting the analysis used may not be an ideal method for selecting genes for further study.

Differential networking proposes that co-expression of genes within disease groups is more important than differences in expression between groups. By performing pair-wise comparisons for co-expression it is possible to identify regulatory relationships between

genes and, eventually, extrapolate whole disease networks [297] and regulatory systems of disease. It would be interesting to carry out this analysis on the current L100P dataset to compare the two methods, and determine whether the networks generated could also be related to human disease.

The evidence for developmental disregulation of candidate genes in the L100P mutant mouse is intriguing, and further work to determine a clear developmental course for the phenotype is highly recommended. Hikida *et al* (2007) [193] previously showed a three month developmental delay in neuronal migration in a dominant-negative *Disc1* mouse model. Kamiya et al (2005) [133] have also shown that expression of dominant-negative mutant *Disc1* in the mouse leads to a delay in neuronal migration at P2, with a reduction in correctly oriented pyramidal cells in the cerebral cortex at P14. Neither experiment looked at the whole developmental time course, but selected key points in postnatal development. Experiments of this type on the L100P ENU mutant mouse would allow a direct comparison between the rate and success of neuronal migration in the developing brain, and the expression levels of *Disc1* and differentially expressed genes. This would allow conclusions to be made as to the effect of gene expression levels on brain development and would provide insight into the development of the 'schizophrenia-like' phenotype in this model.

Once the developmental neuroanatomy of the L100P *Disc1* mouse is established, it would be interesting to see if any deficits can be rescued with the use of drug treatment, and if this also rescues the behavioural phenotype. For example, the *Fmr1* knockout mouse, a model of Fragile-X syndrome – the most common genetic cause of childhood cognitive impairment- shows an increase in dendritic spine maturation and improved behavioural performance when treated with minocyline, a tetracycline analogue [298] from an early age. Minocyline has also been found to have antipsychotic effects, and has been used to help treatment resistance schizophrenia [299]. It would be interesting to test the effect of this, and the other drugs used in this study, in early postnatal animals and adults. This would help us to determine to effect of treatment on behaviour and neuroanatomy, and whether early intervention can rescue both aspects of the phenotype.

8.4 Final comments

It is very clear from this and previous work that the genetics of major mental illness are incredibly complex. The *Disc1* mutant mouse lines have opened the door for further studies into genetic variations in the *DISC1* locus culminating in susceptibility to major mental illness. By increasing our knowledge of the knock-on effects of *Disc1* mutations on down-stream gene functions and affected pathways we could be a step closer to determining the biological mechanism conferring common risk to major mental disorders. A better understanding of the aetiology of schizophrenia, bipolar disorder and other related conditions is crucial for developing rational treatments for the future.

Gene	Probe type	Supplier	Catalogue number	Concentration	Secondary
Egr4	Primary Antibody	Abcam	ab50636	1:16000	Pig anti Rabbit
Nrxn1	Primary Antibody	Abcam	ab79587	.1:100	Rabbit anti Goat
Nrxn3	Primary Antibody	R and D systems	AF5269	.1:100	Donkey anti sheep
Pak3	Primary Antibody	Abcam	ab40808	.1:250	Pig anti Rabbit
Sort1	Primary Antibody	BD Biosciences	612100	.1:500	Rabbit anti Mouse
1500015O10RIK	Taqman qRT-PCR	Applied biosystems	Mm00470447_m1	-	-
Apoe	Taqman qRT-PCR	Applied biosystems	Mm00437573_m1	-	-
Arc	Taqman qRT-PCR	Applied biosystems	Mm01204954_g1	-	-
Atp5b	Taqman qRT-PCR	Applied biosystems	Mm00443967_g1	-	-
Bex2	Taqman qRT-PCR	Applied biosystems	Mm02528127_s1	-	-
Cab39L	Taqman qRT-PCR	Applied biosystems	Mm00471831_m1	-	-
Camk2a	Taqman qRT-PCR	Applied biosystems	Mm00618054_m1	-	-
Cdh11	Taqman qRT-PCR	Applied biosystems	Mm00515458_m1	-	-
Coro1b	Taqman qRT-PCR	Applied biosystems	Mm01231295_gH	-	-
Cplx	Taqman qRT-PCR	Applied biosystems	Mm01349422_g1	-	-
Cstf1	Taqman qRT-PCR	Applied biosystems	Mm00481138_m1	-	-
Disc1	Taqman qRT-PCR	Applied biosystems	Mm00533313_m1	-	-
Dlg2	Taqman qRT-PCR	Applied biosystems	Mm00457160_m1	-	-
Dmrt2	Taqman qRT-PCR	Applied biosystems	Mm01204323_m1	-	-
Dusp1	Taqman qRT-PCR	Applied biosystems	Mm01309843_g1	-	-
Dusp6	Taqman qRT-PCR	Applied biosystems	Mm00518185_m1	-	-
Egr2	Taqman qRT-PCR	Applied biosystems	Mm00456650_m1	-	-
Egr3	Taqman qRT-PCR	Applied biosystems	Mm00516979_m1	-	-
Egr4	Taqman qRT-PCR	Applied biosystems	Mm00842279_g1	-	-
Emid2	Taqman qRT-PCR	Applied biosystems	Mm00502829_m1	-	-
Exoc4	Taqman qRT-PCR	Applied biosystems	Mm00486020_m1	-	-
Fbxo41	Taqman qRT-PCR	Applied biosystems	Mm01189819_g1	-	-
Gad1	Taqman qRT-PCR	Applied biosystems	Mm00725661_s1	-	-
Gapdh	Taqman qRT-PCR	Applied biosystems	Mm99999915_g1	-	-
Gp1bb	Taqman qRT-PCR	Applied biosystems	Mm00494713_s1	-	-
Gpr88	Taqman qRT-PCR	Applied biosystems	Mm00517707_m1	-	-
Grb10	Taqman qRT-PCR	Applied biosystems	Mm00525091_m1	-	-
Gria1	Taqman qRT-PCR	Applied biosystems	Mm00433753_m1	-	-
Gria2	Taqman qRT-PCR	Applied biosystems	Mm00442822_m1	-	-
Gstm1	Taqman qRT-PCR	Applied biosystems	Mm00833915_g1	-	-
Hook3	Taqman qRT-PCR	Applied biosystems	Mm01212616_m1	-	-
Hprt1	Taqman qRT-PCR	Applied biosystems	Mm03024075_m1	-	-
Id1	Taqman qRT-PCR	Applied biosystems	Mm00775963_g1	-	-
Litaf	Taqman qRT-PCR	Applied biosystems	Mm00772072_m1	-	-
Mast3	Taqman qRT-PCR	Applied biosystems	Mm01343218_g1	-	-
Mgfe8	Taqman qRT-PCR	Applied biosystems	Mm0500549_m1	-	-
Mll5	Taqman qRT-PCR	Applied biosystems	Mm01129479_m1	-	-
Mxra7	Taqman qRT-PCR	Applied biosystems	Mm01252810_m1	-	-
Napb	Taqman qRT-PCR	Applied biosystems	Mm00658947_m1	-	-
Ndfip	Taqman qRT-PCR	Applied biosystems	Mm01258332_m1	-	-
Ndn	Taqman qRT-PCR	Applied biosystems	Mm02524479_s1	-	-
Nrgn	Taqman qRT-PCR	Applied biosystems	Mm00480741_m1	-	-
Nrxn1	Taqman qRT-PCR	Applied biosystems	Mm00660298_m1	-	-
Nrxn3	Taqman qRT-PCR	Applied biosystems	Mm01335648_m1	-	-
Numb	Taqman qRT-PCR	Applied biosystems	Mm01302750_m1	-	-
Pak3	Taqman qRT-PCR	Applied biosystems	Mm00435482_m1	-	-
Pcb1	Taqman qRT-PCR	Applied biosystems	Mm00478712_s1	-	-
Pde1a	Taqman qRT-PCR	Applied biosystems	Mm01275734_m1	-	-
Pde4d	Taqman qRT-PCR	Applied biosystems	Mm01304777_m1	-	-
Plk1	Taqman qRT-PCR	Applied biosystems	Mm00440924_g1	-	-
Plxnb2	Taqman qRT-PCR	Applied biosystems	Mm00507118_m1	-	-
Pop4	Taqman qRT-PCR	Applied biosystems	Mm00546481_m1	-	-
Ppargc1	Taqman qRT-PCR	Applied biosystems	Mm01208835_m1	-	-
Prm3	Taqman qRT-PCR	Applied biosystems	Mm00443095_s1	-	-
Prmt6	Taqman qRT-PCR	Applied biosystems	Mm01206465_s1	-	-
Psmd1	Taqman qRT-PCR	Applied biosystems	Mm01295362_m1	-	-
Ranbp9	Taqman qRT-PCR	Applied biosystems	Mm00451306_m1	-	-
Rapgef5	Taqman qRT-PCR	Applied biosystems	Mm00616225_m1	-	-
s100a8	Taqman qRT-PCR	Applied biosystems	Mm00496696_g1	-	-
s100a9	Taqman qRT-PCR	Applied biosystems	Mm00656925_m1	-	-
Shank3	Taqman qRT-PCR	Applied biosystems	Mm00498775_m1	-	-
Sort1	Taqman qRT-PCR	Applied biosystems	Mm00490905_m1	-	-
Syn1	Taqman qRT-PCR	Applied biosystems	Mm00449772_m1	-	-
Syn2	Taqman qRT-PCR	Applied biosystems	Mm00449780_m1	-	-
Syp	Taqman qRT-PCR	Applied biosystems	Mm00436850_m1	-	-
Tgfbi	Taqman qRT-PCR	Applied biosystems	Mm00493634_m1	-	-
Ubl3	Taqman qRT-PCR	Applied biosystems	Mm00449451_m1	-	-
Xbp1	Taqman qRT-PCR	Applied biosystems	Mm00457360_g1	-	-
Wdfy1	Taqman qRT-PCR	Applied biosystems	Mm00840455_m1	-	-

Probe information for qRT-PCR and protein analysis. Column 1 gives gene target ID, Column 2 probe type, Column 3 supplier, Column 4 catalogue number, Column 5 concentration (where applicable) and Column 6 Secondary antibody used(were applicable)

References

1. Murray, CJL. *The Global Burden of disease: A comprehensive assessment of mortality, injuries and risk factors in 1990 and projected to 2020.* . in *Harvard School of Public Health and the World Health Organisation* 1996. Boston.
2. Berner, P., Gabriel, E, Katschnig, H, Kieffer, W, Koehler, K, Lenz, G, *Diagnostic criteria for functional psychoses 2nd ed.* 1992, Cambridge: Cambridge University Press.
3. Owen, M.Craddock, N O'Donovan, MC *Schizophrenia: Genes at last?*. . Trends in Genetics 2005(21): p. 518-525.
4. Swann, A.C., *Bipolar disorder and schizophrenia: Continuum or distinct illnesses?* 2nd Biennial Conference of the International Society for Bipolar Disorders, 2006.
5. Lichtenstein P, Yip BH, Bjork C, Pawitan Y, Cannon TD, Sullivan PF, Hultman CM, *Common genetic determinants of schizophrenia and bipolar disprder in swedish families: a population based study.* Lancet, 2009. **373**(9659): p. 234-239.
6. Gottesman II, Laursen TM, Bertelsen A, Mortensen PB, *Severe mental disorders in offspring with 2 psychiatrically ill parents.* Archives of General Psychiatry, 2010. **67**(3): p. 252-257.
7. Purcell SM, Wray N, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P, ISC, *Common polygenic variation contributes to risk of schizophrenia and bipolar disorder.* Nature, 2009. **460**(7256): p. 748-752.
8. Blackwood, D., Fordyce, A, Walker, MT, St. Clair, DM, Porteous, DJ, Muir, WJ, *Schizophrenia and Affective Disorders—Cosegregation with a Translocation at Chromosome 1q42 That Directly Disrupts Brain-Expressed Genes: Clinical and P300 Findings in a Family.* American Journal of Human Genetics, 2001. **69**: p. 428-433.
9. *Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR, Fourth Edition, Text Revision*. 2000, American Psychiatric Association: Washington DC.
10. Siris, S., *Suicide and Schizophrenia.* Journal of Psychopharmacology, 2001. **15**(2): p. 127-135.
11. Rodgers, T., *NICE recommends newer antipsychotic drugs as one of the first line options for schizophrenia*. 2002, NICE Press Release.
12. Angermeyer, M and Kuhn L, *Gender differences in age at onset of schizophrenia.* . Eur Arch Psychiatry and Neuro Sciences 1988(237): p. 351-364.
13. Lewine R, *At issue: Sex and Gender in schizophrenia.* Schizophrenia Bulletin, 2004. **30**(4): p. 755-762.
14. Bardenstein KK, McGlashan T., *Gender differences in affective, schizoaffective and schizophrenic disorders. A review.* Schizophrenia research, 1990. **3**(3): p. 159-172.
15. Leung A, Chue P., *Sex differences in schizophrenia. A review of the literature.* Acta Psychiatr Scand, 2000. **401**: p. 3-38.
16. Hafner H, *Gender differences in schizophrenia*. Psychoneuroendochrinology, 2003. **Suppl2**: p. 17-54.
17. Breier A, Buchannan R., Elkashef A, Munson RC, Kirkpatric B, Gellad F, *Brain morphology and schizophrenia. A magnetic resonance imaging study of limbic, prefrontal cortex and caudate structures.* Arch Gen Psychiatry, 1992. **49**(12): p. 921-926.

18. Lieberman J, Chakos M., Huwei W, Alvir J, Hoffman E, Robertson D and Bilder R *Longitudinal study of brain morphology in first episode schizophrenia,*. Biological Psychiatry 2001 (49): p. 487-499.
19. Ellison-Wright I, Bullmore E., *Anatomy of bipolar disorder and schizophrenia: a meta-analysis.* Schizophr Res. , 2010. **117**(1): p. 1-12.
20. Wright IC, Rabe-Hesketh S., Woodruff PWR, David AS, Murray RM, Bullmore ET, *Meta-Analysis of Regional Brain Volumes in Schizophrenia*. Am J Psychiatry, 2000. **157**: p. 16-25.
21. Adriano F, Spoletini I., Caltagirone C, Spalletta G, *Updated meta-analyses reveal thalamus volume reduction in patients with first-episode and chronic schizophrenia.* Schizophr Res., 2010. **123**(1): p. 1-14.
22. Di X, Chan R., Gong QY., *White matter reduction in patients with schizophrenia as revealed by voxel-based morphometry: an activation likelihood estimation meta-analysis.* Prog Neuropsychopharmacol Biol Psychiatry, 2009. **33**(8): p. 1390-1394.
23. Keshavana MS, Diwadkar V., Montrose DM, Rajarethinama R, Sweeney JA, *Premorbid indicators and risk for schizophrenia: A selective review and update* Schizophr Res., 2005. **79**(1): p. 45-57.
24. Naheed M and Green B. *Focus on Clozapine*. 2000 [cited; Available from: http://www.priory.com/focus14.htm.
25. Naheed M, and Green B., *Focus on Clozapine.* Curr Med Res Opin, 2001. **17**(3): p. 223-229.
26. Citrome L, Builder R., Volavka J, *Managing Treatment-Resistant Schizophrenia: Evidence from Randomized Clinical Trials.* Journal of Psychiatric Practice 2002. **8**: p. 205–215.
27. Alvir JMA, Lieberman J., Safferman AZ, Schwimmer JL, Schaaf JA *Clozapine-Induced Agranulocytosis -- Incidence and Risk Factors in the United States* New England Journal of Medicine 1993. **329**: p. 162-167.
28. Conley RR, Kelly D., *Current status of antipsychotic treatment.* . Curr Drug Targets CNS Neurol Disord 2002. **1**: p. 123–128.
29. Awad AG, Voruganti L., *New antipsychotics, compliance, quality of life, and subjective tolerability. Are patients better off?* . Can J Psychiatry 2004. **49**: p. 297–302.
30. Weickert TW, Goldberg T., *First- and second-generation antipsychotic medication and cognitive processing in schizophrenia.* Curr Psychiatry Rep 2005. **7**: p. 304–310.
31. Hennekens CH, Hennekens AR, Hollar D, Casey DE *Schizophrenia and increased risks of cardiovascular disease.* . American Heart Journal 2005. **150**: p. 1115–1121.
32. Seeman, M., *An outcome measure in schizophrenia: mortality.* Can J Psychiatry 2007. **52**: p. 55–60.
33. Auquier P, Lancon C, Rouillon F, Lader M, Holmes C *Mortality in schizophrenia.* . Pharmacoepidemiol Drug Saf 2006. **15** p. 873–879.
34. Coyle JT, Balu D., Benneyworth M, Basu A, Roseman A., *Beyond the dopamine receptor: novel therapeutic targets for treating schizophrenia.* Dialogues of Clinical Neuroscience, 2010. **12**(3): p. 359-382.

35. Kulkarni J, de Castella A., Headley B, Marston N, Sinclair K, Lee S, Gurvich C, Fitzgerald PB, Burger H, *Estrogens and men with schizophrenia: Is there a case for adjunctive therapy?* Schizophr Res., 2010. **in press**.
36. Kane JM, Correll C., *Pharmacologic treatment of schizophrenia.*. Dialogues of Clinical Neuroscience, 2010. **12**(3): p. 345-357.
37. Dumitriu D, Rapp P., McEwen BS, Morrison JH., *Estrogen and the aging brain: an elixir for the weary cortical network.* Ann NY Acad Sci, 2010. **1204**: p. 104-112.
38. Seeman MV, Lang M., *The role of estrogens in schizophrenia gender differences.* Schizophrenia Bulletin, 1990. **16**(2): p. 185-194.
39. Goodwin FK, Jamison K., *Manic depressive illness* 1990, New York: Oxford University press
40. Chen, Y.-W, Dilsaver, SC, *Lifetime rates of suicide attempts among subjects with bipolar and unipolar disorders relative to subjects with other Axis I disorders.* Biol. Psychiatry, 1996. **39**: p. 896–899.
41. Das Gupta R, Guest J., *Annual Cost of bipolar disorder to UK society*. Br. J. Psychiatry, 2002. **180**: p. 227-233.
42. Kawa, I., Carter, JD, Joyce, PR, Doughty, CJ, Frampton, CM, Wells, JE, Walsh, AES, Olds, RJ, *Gender differences in bipolar disorder: age of onset, course, comorbidity, and symptom presentation.* Bipolar Disorders, 2005. **7**(2): p. 119-125.
43. Nivoli AM, Murru A., Vieta E, *Lithium: Still a cornerstone in the longterm treatment of bipolar disorder?* Neuropsychobiology, 2010. **62**(1): p. 27-35.
44. NONAKA, S., HOUGH, CJ, CHUANG, D, *Chronic lithium treatment robustly protects neurons in the central nervous system against excitotoxicity by inhibiting N-methyl-D-aspartate receptor-mediated calcium influx.* Proc. Natl. Acad. Sci. USA, 1998. **95**: p. 2642–2647.
45. Ustun TB, Kessler R., *Global Burden of depressive disorders: the issue of duration.* Br. J. Psychiatry, 2002. **181**: p. 181-183.
46. Bollini P, Pampallona S., Tibaldi G, Kupelnick B, Munizza C, *Effectiveness of antidepressants. Meta-analysis of dose-effect relationships in randomised clinical trials.* Br. J. Psychiatry, 1999. **174**: p. 297-303.
47. Nierenberg AA, Keefe B., Leslie VC, Alpert JE, Pava JA, Worthington JJ, Rosenbaum JF, Fava M, *Residual symptoms in depressed patients who respond acutely to fluoxetine.* J Clinical Psychiatry, 1999. **60**(4): p. 221-225.
48. Tranter R, O'Donovan C., Chandarana P, Kennedy S, *Prevelance and outcome of partial remission in depression.* J Psychiatry Neuroscience, 2002. **27**(4): p. 241-247.
49. Thaker, G.K.and Carpenter W.T *Advances in schizophrenia.* . Nature Medicine 2001(7): p. 667-671.
50. Cannon, T., Huttunen, MO, Lonnqvist, J, Tuulio-Henriksson, A, Pirkola, T, Glahn, D, Finkelstein, J, Hietanen, M, Kaprio, J, Koskenvuo, M, *The Inheritance of Neuropsychological Dysfunction in Twins Discordant for Schizophrenia* American Journal of Human Genetics, 2000(67): p. 369-382.
51. Prescott, C.A and Gottesman I.I *Genetically mediated vulnerability to schizophrenia.* . Psychiatric Clin North Am, 1993(16): p. 245-267.

52. Risch, N., *Linkage Strategies for Genetically complex traits I. Multilocus models.* American Journal of Human Genetics, 1990(46): p. 222-228.
53. Kieseppa T, Partonen T., Haukka J, Kaprio J and Lonnqvist J, *High concordance of bipolar I disorder in a nationwide sample of twins.* American Journal of Psychiatry, 2004(161): p. 1814-1821.
54. Dick, D., Foroud, T, Flury, L, Bowman, ES, Miller, MJ, Rau, NL, Moe, PR, Samavedy, N, El-Mallakh, R, Manji, H,, D. Glitz, Meyer, ET, Smiley, C, Hahn, R, Widmark, C,, and R. McKinney, Sutton, L, Ballas, C, Grice, D, Berrettini,W, Byerley, W, Coryell, W, DePaulo, R, MacKinnon, DF, Gershon, ES, Kelsoe, JR, McMahon, FJ, McInnis, M, Murphy, DL, Reich, T, Scheftner, W, Nurnberger Jr,JI, *Genomewide Linkage Analyses of Bipolar Disorder: A New Sample of 250 Pedigrees from the National Institute of Mental Health Genetics Initiative.* American Journal of Human Genetics, 2003(73): p. 107-114.
55. Maier W, Rietschel M., Lichtermann D, Wildenauer D, *Family and genetic studies on the relationship of schizophrenia to affective disorders.* European Archives of Psychiatry and Clinical neuroscience, 1999. **Supp 4**(249): p. 57-61.
56. Millar, J., Wilson-Annan, JC, Anderson, S, Christie, S, Taylor, MS, Semple, CAM, Devon, RS, St Clair, DM, Muir, WJ, Blackwood, DHR, Porteous, DJ, *Disruption of two novel genes by a translocation co-segregating with schizophrenia.* Human Molecular Genetics, 2000. **9**(9): p. 1415-1423.
57. Venken T, Del-Favero J., *Chasing genes for mood disorders and schizophrenia in genetically isolated populations.* Human Mutat, 2007. **28**(12): p. 1156-1170.
58. Craddock N, O'Donovan M., Owen MJ *The genetics of schizophrenia and bipolar disorder: dissecting psychosis. .* Journal Medical Genetics 2005(42): p. 193-204.
59. Serretti A, Mandelli L., *The genetics of bipolar disorder: genome 'hot' regions, genes, new potential candidates and future directions.* Mol Psychiatry, 2008. **13**(8): p. 742-771.
60. Alaerts M, Del-Favero J., *Searching genetic risk factors for schizophrenia and bipolar disorder: learn from the past and back to the future.* Human Mutat, 2009. **30**(8): p. 1139-1152.
61. Jeffreys AJ, Kauppi L., Neumann R, *Intensely punctate meiotic recombination in the classII region of the major hiatocompatibility complex.* Nat Genet, 2001. **29**(2): p. 217-222.
62. Wall JD, Pritchard J., *Assessing the performance of the haplotype block model of linkage disequilibrium.* Am J Human Genetics, 2003. **73**(3): p. 502-515.
63. Goldstein DB, Ahmandi K., Weale ME, Wood NW, *Genome scans and candidate gene approaches in the study of common diseases and variable drug responses.* Trends in Genetics, 2003. **19**(11): p. 615-622.
64. Hennah W, Thomson P, Peltonen L and Porteous D *Beyond schizophrenia: The role of DISC1 in Major Mental Illness, .* Schizophrenia Bulletin, 2006. **3**(32): p. 409-416.
65. Chubb JE, Bradshaw N., Soares DC, Porteous DJ, Millar JK, *The DISC locus in psychiatric illness.* Molecular Psychiatry, 2008. **13**: p. 36-64.
66. Stefansson H, Sarginson J., Kong A, Yates P, Steinthorsdottir V, Gudfinnsson E, Gunnardottir S, Walker N, Petursson H, Crombie C, Ingason A, Gulcher JR,

Stefansson K, St Clair D, *Association of neuregulin 1 with schizophrenia confirmed in a scottish population.* Am J Human Genetics, 2003. **72**(1): p. 83-87.
67. Sullivan, P., *The genetics of schizophrenia.* PLoS medicine, 2005. **2**(7): p. 614-618.
68. Kitsios GD, Zintzaras E., *Genome-wide association studies: hypothesis-'free' or 'engaged'*. Transl Res, 2009. **154**(4): p. 161-164.
69. Stefansson H, Ophoff R., Steinberg S, Andreassen OA, Cichon S, Rujescu D, Werge T, Collier DA et al, *Common variants conferring risk of schizophrenia.* Nature, 2009. **460**(7256): p. 744-747.
70. Kirov G, Zaharieva I., Georgieva L, Moskvina V, Nikolov I, Cichon S, Hilmer A, Toncheva D, Owen MJ, O'Donovan MC, *A genome-wide association study in 574 schizophrenia trios using DNA pooling.* Mol Psychiatry, 2009. **14**(8): p. 796-803.
71. Kanarky CG, Li Z., Nakai Y, Sei Y, Weinberger DR, *Neuregulin-1 regulates cell adhesion via an ErbB2/phosphoinositide-3 kinase/AKT dependent pathway:potential implications for schizophrenia and cancer*. PloS One, 2007. **2**(12): p. e1369.
72. Shifman S, Johannesson M., Bronstein M, Chen SX, Collier DA, Craddock NJ, Kendler KS, Li T, O'Donovan M, O'Neill FA, Owen MJ, Walsh D, Weinberger DR, Sun C, Flint J, Darvasi A, *Genome-wide association identifies a common varient in the reelin gene that increase the risk of schizophrenia only in women.* PLoS Genetics, 2008. **4**(2): p. e28.
73. Sun S, Wang F., Wei J, Cao LY, Wu GY, Lu L, Kosten TA, Kosten TR, Zhang XY, *Association between interleukin-3 receptor alpha polymorphism and schizophrenia in the chinese population.* Neurosci Lett., 2008. **440**(1): p. 35-37.
74. Loe-Mie Y, Lepagnol-Bestel A., Maussion G, Doron-Faigenboim A, Imbeaud S, Delacroix H, Aggerbeck L, Pupko T, Gorwood P, Simmonneau M, Moalic JM, *SMARC2 and other genome wide supported schizophrenia associated genes: regulation by REST/NRSF, network organisation and primate specific evolution* Hum Mol Genet, 2010. **19**(14): p. 284-257.
75. Athanasiu L, Mattingsdak M., Kahler AK, Brown A, Gustafsson O, Agartz I, Giegling I, Muglia P, Cichon S, Rietschel M, Pietilainen OP, Peltonen L, Bramon E, Collier D, Clair DS, Sigurdsson E, Petursson H, Rujescu D, Melle I, Steen VM, Djurovic S, Andreassen OA, *Gene variants associated with schizophrenia in a Norwegian genome-wide study are replicated in a large european cohort.* J Psychiatric research, 2010. **Epub ahead of print**.
76. The Wellcome Trust Case Control Consortium, *Genome-wide association study of 14000 cases of seven common diseases and 3000 shared controls.* Nature, 2007. **447**: p. 661-678.
77. Sullivan, P., *The Psychiatric GWAS Consortium: Big science comes to Psychiatry.* Neuron, 2010. **68**: p. 182-186.
78. Williams HJ, Norton N, Dwyer S, Moskvina V, Nikolov I, Carroll L, Georgieva L, Williams NM, Morris DW, Quinn EM, Giegling I, Ikeda M, Wood J, Lencz T, Hultman C, Lichtenstein P, Thiselton D, Maher BS, Molecular Genetics of Schizophrenia Collaboration (MGS), International Schizophrenia Consortium (ISC), SGENE-plus, Malhotra AK, Riley B, Kendler KS, Gill M, Sullivan P, Sklar P, Purcell S, Nimgaonkar VL, Kirov G, Holmans P, Corvin A, Rujescu D,

Craddock N, Owen MJ and O'Donovan MC, *Fine mapping of ZNF804A and genome-wide significant evidence for its involvement in schizophrenia and bipolar disorder.* Molecular Psychiatry, 2010: p. 1-13.

79. O'Donovan MC, Norton N., Williams H, Peirce T, Moskvina V, Nikolov I, Hamshere M, Carroll L, Georgieva L, Dwyer S, Holmans P, Marchini JL, Spencer CCA, Howie B, Leung H-T, Giegling I, Hartmann AM, Mo¨ ller H-J, Morris DW, Shi Y, Feng G, Hoffmann P, Propping P, Vasilescu C, Maier W, Rietschel M, S Zammit S, Schumacher J, Quinn EM, Schulze TG, Iwata N, Ikeda M, Darvasi A, Shifman S, He L, Duan J, Sanders AR, Levinson DF, Adolfsson R, O¨ sby U, Terenius L, Jo¨nsson EG, Cichon S, No¨then MM, Gill M, Corvin AP, Rujescu D, Gejman PV, Kirov G, Craddock N, Williams NM and Owen MJ, Molecular Genetics of Schizophrenia Collaboration, *Analysis of 10 independent samples provides evidence for association between schizophrenia and a SNP flanking fibroblast growth factor receptor 2.* Molecular Psychiatry, 2009. **14**: p. 30-36.
80. Caspi A, Sugden K., Moffitt T.E, Taylor A, Craig I.W Harrington H, McClay J, Mill J, Martin J, Braithwaite A and Poulton R *Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene* Science 2003 (301): p. 386-389.
81. Craddock N, O'Donovan M.C., Owen M.J *The genetics of schizophrenia and bipolar disorder: dissecting psychosis. .* Journal Medical Genetics 2005(42): p. 193-204.
82. Kirov , G., *The role of copy number variation in schizophrenia.* Expert review of Neurotheraputics, 2010. **10**(1): p. 25-32.
83. Walsh T, McClellan J., McCarthy SE, Addington AM, Pierce SB, Cooper GM, Nord AS, Kusenda M, Malhotra D, Bhandari A *Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia.* Science, 2008. **320**: p. 539-543.
84. Xu B, Roos J., Levy S, van Rensburg EJ, Gogos JA and Karayiorgou M *Strong association of de novo copy number mutations with sporadic schizophrenia.* Nature Genetics, 2008. **40**: p. 880-885.
85. Saus E, Brunet A., Armengol L, Alosno P, Crespo JM, Fernandez-Aranda F, Guitart M, Martin-Santos R, Menchon JM, Navines R, Soria V, Torrens M, Urretavizcaya M, Valles V, Gratacos M, Estivill X, *Comprehensive copy number varient (CNV) analysis of neuronal pathway genes in psychiatric disorders identifies rare variants within patients.* J Psychiatric research, 2010. **Epub ahead of print**.
86. Need AC, Ge D., Weale ME, Maia J, Feng S, Heinzen EL, Shianna KV, Yoon W, Kasperavic˘iu¯ te˙ D, Gennarelli M, Strittmatter WJ, Bonvicini C, Rossi G, Jayathilake K, Cola PA, McEvoy JP, Keefe RSE, Fisher EMC, St. Jean PL, Giegling I, Hartmann AM, Mo¨ ller H-J, Ruppert A, Fraser G, Crombie C, Middleton LT, St. Clair D, Roses AD, Muglia P, Francks C, Rujescu D, Meltzer HY, Goldstein DB, *A Genome-Wide Investigation of SNPs and CNVs in Schizophrenia.* PLoS Genetics, 2009. **5**(2).
87. Sawa A, Snyder S., *Schizophrenia: diverse approaches to a complex disease.* Science, 2002. **296**: p. 692-695.

88. Mirnics K, Middleton F., Marques A, Lewis DA, Levitt P, *Molecular characterisation of schizophrenia viewed by microarray analysis of gene expression in prefrontal cortex.* Neuron, 2000. **28**: p. 53-67.
89. Vawter MP, Crook J., Hyde TM, Kleinman JE, Weinberger DR, Becker KG, Freed WJ, *Microarray analysis of gene expression in the prefrontal cortext in schizophrenia: a preliminary study.* Schizophrenia research, 2002. **58**: p. 11-20.
90. Collinge, J., and Curtis, D. , *Decreased hippocampal expression of a glutamate receptor gene in schizophrenia.* . Br. J. Psychiatry, 1991. **159**: p. 857–859.
91. Volk, D.V., Austin, M.C., Pierri, J.N., Sampson, R.A., and Lewis, D.A., *Decreased GAD67 mRNA expression in a subset of prefrontal cortical GABA neurons in schizophrenia.* . Arch. Gen. Psychiatry, 2000. **57**: p. 237-245.
92. SUGAI T, K.M., IRITANI S, ARAKI K, MAKIFUCHI T, IMAI C, NAKAMURA R, and T.H. KAKITA A, NAWA H, *Prefrontal Abnormality of Schizophrenia Revealed by DNA Microarray Impact on Glial and Neurotrophic Gene Expression.* Ann. N.Y. Acad. Sci. , 2004. **1025:** : p. 84–91.
93. Bowden NA, Scott R., Tooney PA, *Altered gene expression in the superior temporal gyrus in schizophrenia.* BMC Genomics. , 2008. **9**.
94. Iwamoto K, Bundo M., Kato T., *Altered expression of mitochondria-related genes in postmortem brains of patients with bipolar disorder or schizophrenia, as revealed by large-scale DNA microarray analysis.* Hum Mol Genet., 2005. **14**(2): p. 241-53.
95. Prabakaran S, Swatton J., Ryan MM, Huffaker SJ, Huang JT, Griffin JL, Wayland M, Freeman T, Dudbridge F, Lilley KS, Karp NA, Hester S, Tkachev D, Mimmack ML, Yolken RH, Webster MJ, Torrey EF, Bahn S., *Mitochondrial dysfunction in schizophrenia: evidence for compromised brain metabolism and oxidative stress.* Mol Psychiatry., 2004 **9**(7): p. 684-697.
96. Bunney WE, Bunney B., Vawter MP, Tomita H, Li J, Evans SJ, Choudary PV, Myers RM, Jones EG, Watson SJ, Akil H, *Microarray Technology: A Review of New Strategies to Discover Candidate Vulnerability Genes in Psychiatric Disorders* Am J Psychiatry, 2003. **160**: p. 657–666.
97. Glatt SJ, Everall I., Kremen WS, Corbeil J, ˇSa´ sˇik R, Khanlou N, Han M, Liew C and Tsuang MT, *Comparative gene expression analysis of blood and brain provides concurrent validation of SELENBP1 up-regulation in schizophrenia* PNAS, 2005. **102**(43): p. 15533-15538.
98. Lake, R., *Disorders of Thought Are Severe Mood Disorders: the Selective Attention Defect in Mania Challenges the Kraepelinian Dichotomy—A Review.* Scizophrenia Bulletin Advance Access, 2007.
99. Mirnics, K., Levitt, P, and Lewis, DA, *Critical Appraisal of DNA Microarrays in Psychiatric Genomics.* Biological Psychiatry, 2006. **60**: p. 163-176.
100. Pickard BS, Millar JK, Porteous DJ, Muir WJ, Blackwood DHR, *Cytogenetics and gene discovery in psychiatric disorders.* The Pharmacogenomics journal, 2005. **5**: p. 81-88.
101. Jacobs PA, Brunton M., Frackiewicz A, Newton M, Cook PJL, Robson EB *Studies on a family with three cytogenetic markers.* Ann Hum Genet (Lond), 1970. **33**: p. 325–336.

102. St Clair D, Blackwood D., Muir W, Carothers A, Walker M, Spowart G, Godsen C and Evans H.J *Association within a family of a balanced autosomal translocation with major mental illness,* . The Lancet 1990 (336): p. 13-16.
103. Millar J.K, James R., Brandon N and Thomson P.A *DISC1 and DISC2: discovering and dissecting molecular mechanisms underlying psychiatric illness.* . Trends in Molecular Medicine, 2004.
104. Millar, J.K., et al., *Disrupted in schizophrenia 1 (DISC1): Subcellular targeting and induction of ring mitochondria.* Molecular and Cellular Neuroscience, 2005. **30**(4): p. 477-484.
105. Zhou X, Chen Q., Schaukowitch K, Kelsoe JR, Geye MA, *Insoluble DISC1-Boymaw fusion proteins generated by DISC1 translocation.* Molecular Psychiatry, 2010. **15**: p. 670-672.
106. Brandon, N., Handford, EJ, Schurov, I, Rain, J-C, Pelling, M, Duran-Jimeniz, B, Camargo, LM, Oliver, KR, Beher, D, Shearman, MS, Whiting, PJ, *Disrupted in Schizophrenia 1 and Nudel form a neurodevelopmentally regulated protein complex: implications for schizophrenia and other major neurological disorders.* Molecular and cellular neuroscience, 2004. **25**: p. 42-55.
107. Hennah W, Thomson P., Peltonen L and Porteous D *Beyond schizophrenia: The role of DISC1 in Major Mental Illness,* . Schizophrenia Bulletin 2006. **3**(32): p. 409-416.
108. Maeda, K., Nwulia, E, Chang, J, Balkissoon, R, Ishizuka, K, Chen, H, Zandi, P, McInnis, MG, Akira Sawa, A, *Differential Expression of Disrupted-in-Schizophrenia (DISC1) in Bipolar Disorder.* Biological Psychiatry, 2006. **60**: p. 929-935.
109. Hodgkinson, C., Goldman, D, Jaeger, J, Persaud, S, Kane, JM, Lipsky, RH, Malhotra, AK, *Disrupted in Schizophrenia 1 (DISC1): Association with Schizophrenia, Schizoaffective Disorder, and Bipolar Disorder* American Journal of Human Genetics, 2004. **75**: p. 862-872.
110. Sachs, N., Sawa, A, Holmes, SE, Ross, CA, DeLisi, LE, Margolis, RL, *A frameshift mutation in Disrupted in Schizophrenia 1 in an American family with schizophrenia and schizoaffective disorder.* Molecular Psychiatry, 2005. **10**: p. 758-764.
111. Kockelhorn T.T, Arai M., Matsumoto H, Fukuda N, Yamada K, Minabe Y, Toyota T, Ujike H, Sora I, Mori N, Yoshikawa T and Hokawa M *Association study of polymorphisms in the 5` upstream region of human DISC1 gene with schizophrenia.* . Neurosciende letters 2004(368): p. 41-45.
112. Devon R.S, Anderson S., Teague P.W, Burgess P, Kipari T.M, Semple C.A, Millar J.K, Muir W.J, Murray V, Pelosi A.J, Blackwood D.H and Porteous D.J *Identification of polymorphisms within Disrupted in Schizophrenia 1 and Disrupted in Schizophrenia 2, and an investigation of their association with schizophrenia and bipolar affective disorder.* . Psychiatric Genetics 2001(11): p. 71-78.
113. Zhang X, Tochigi M., Ohashi J, Maeda K, Kato T, Okazaki Y, Kato N, Tokunaga K, Sawa A and Sasaki T Schizophrenia Research 2005(79): p. 175-180.
114. Hennah, W., Varilo, T, Kestila¨, M, Paunio, T, Araja¨ rvi, R, Haukka, J, Parker, A, Martin, R, Levitzky, S, Partonen, T, Meyer, J, Lo¨ nnqvist, J, Peltonen, L,

Ekelund, J, *Haplotype transmission analysis provides evidence of association for DISC1 to schizophrenia and suggests sex-dependent effects.* Human Molecular Genetics, 2003. **12**: p. 3151-3159.
115. Hennah W, Tuulio-Henriksson A., Paunio T, Ekelund J, Varilo T, Partonen T, Cannon TD, Lonnqvist J, Peltonen L, *A haplotype within the DISC1 gene is associated with visual memory functions in families with a high density of schizophrenia.* Mol Psychiatry, 2005. **10**(12): p. 1097-1103.
116. Cannon TD, Hennah W., van Erp TG, Thompson PM, Lonnqvist J, huttunen M, Gasperomi T, Tuulio-Hennriksson A, Pirkola T, Toga AW, Kapiro J, Mazziotta J, Peltonen L, *Association of DISC1/TRAX haplotypes with schizophrenia, reduced prefrontal grey matter and impaired short and long-term memory.* Arch Gen Psychiatry, 2005. **62**(11): p. 1205-1213.
117. Zhang F, Sarginson J., Crombie C, Walker N, St Clair D, Shaw D, *Genetic association between schizophrenia and the disc1 gene in the scottish population.* Am J Med Genet B Neuropsychiatr Genet, 2005. **141B**(2): p. 155-159.
118. Kilpinen H, Ylisaukko-Oja T., Hennah W, Palo OM, Varilo T, Vanhala R, Nieminen-von Wendt T, von Wendt L, Paunio T, Peltonen L., *Association of DISC1 with autism and Asperger syndrome.* Molecular Psychiatry, 2008. **13**(2): p. 187-196.
119. Hamshere ML, Bennett P., Williams N, Segurado R, Cardno A, Norton N, Lambert D, Williams H, kirov G, Corvin A, Holmans P, Jones L, Jones I, Gill M, O'Donovan MC, Owen MJ, Craddock N, *Genomewide linkage scan in schizoaffective disorder: significant evidence for linkage at 1q42 close to DISC1, and suggestive evidence at 22q11 and 19p13.* Arch Gen Psychiatry, 2005. **62**(10): p. 1081-1088.
120. Hodgkinson C.A, Goldman D., Jaeger J, Persuad S, Kane J.M, Lipsky R.H and Malhotra A.K *Disrupted in schizophrenia 1 (DISC1): Association with schizophrenia, schizoaffective disorder and bipolar disorder.* . American Journal of Human Genetics 2004 (75): p. 862-872.
121. Burdick KE, Hodgkinson C., Szeszko PR, Lencz T, Ekholm JM, Kane JM, Goldman D, Malhotra AK., *DISC1 and neurocognitive function in schizophrenia.* . Neuroreport., 2005 **16**(12): p. 1399-1402.
122. Thomson PA, Harris S., Starr JM, Whalley LJ, Porteous DJ, Deary IJ., *Association between genotype at an exonic SNP in DISC1 and normal cognitive aging.* . Neurosci Lett., 2005. **389**(1): p. 41-45.
123. Callicott JH, Straub R., Pezawas L, Egan MF, Mattay VS, Hariri AR, Verchinski BA, Meyer-Lindenberg A, Balkisson R, Kolachana B, Goldberg TE, Weinberger DR, *Variation in DISC1 affects hippocampal structure and function and increases risk for schizophrenia.* Proc Nat Acad Sci USA, 2005. **102**(24): p. 8627-8632.
124. Di Giorgio A, Blasi G., Sambataro F, Rampino A, Papazacharias A, Gambi F, Romano R, Caforio G, Rizzo M, Latorre V, Popolizio T, Kolachana B, Callicott JH, Nardini M, Weinberger DR, Bertolino A., *Association of the SerCys DISC1 polymorphism with human hippocampal formation gray matter and function during memory encoding.* Eur J Neurosci. , 2008. **28**(10): p. 2129-3216.
125. Prata DP, Mechelli A., Fu CH, Picchioni M, Kane F, Kalidindi S, McDonald C, Kravariti E, Toulopoulou T, Miorelli A, Murray R, Collier DA, McGuire PK.,

Effect of disrupted-in-schizophrenia-1 on pre-frontal cortical function. Mol Psychiatry., 2008. **13**(10): p. 915-917.
126. Hashimoto R, Numakawa T., Ohnishi T, Kumamaru E, Yagasaki Y, Ishimoto T, Mori T, Nemoto K, Adachi N, Izumi A, Chiba S, Noguchi H, Suzuki T, Iwata N, Ozaki N, Taguchi T, Kamiya A, Kosuga A, Tatsumi M, Kamijima K, Weinberger DR, Sawa A, Hiroshi H, *Impact of the DISC1 Ser704Cys polymorphism on risk for major depression, brain morphology and ERK signaling* Human Molecular Genetics 2006. **15**(20): p. 3024-3033.
127. Song W, Li W., Feng J, Heston LL, Scaringe WA, Sommer SS, *Identification of high risk DISC1 structural cariants with a 2% attributable risk for schizophrenia.* Biochem Biophys Res Commun, 2008. **367**(3): p. 700-706.
128. Hennah W, and Porteous D., *The DISC1 Pathway Modulates Expression of Neurodevelopmental, Synaptogenic and Sensory Perception Genes.* PloS One, 2009. **4**(3): p. e4906.
129. Eastwood SL, Hodgkinson C., Harrison PJ, *DISC1 Leu607Phe allels differentially affect chromosomal PCM1 localisation and neurotransmitter release.* Mol Psychiatry, 2009. **14**(6): p. 556-557.
130. Nakata K, Lipska B., Hyde TM, Ye T, Newburn EN, Morita Y, Vakkalanka R, Barenboim M, Sei Y, Weinberger DR, Kleinman JE, *DISC1 splice variants are upregulated in schixophrenia and associated with risk polymorphisms.* Proc Nat Acad Sci USA, 2009. **106**(37): p. 15873-15878.
131. Millar, J.K., et al., *DISC1 and PDE4B are interacting genetic factors in schizophrenia that regulate cAMP signaling.* Science, 2005. **310**(5751): p. 1187-91.
132. Camargo, L., Collura, V, Rain, J-C, Mizuguchi, K, Hermjakob,H, Kerrien, S, Bonnert, TP, Whiting, PJ, Brandon, NJ, *Disrupted in Schizophrenia 1 Interactome: evidence for the close connectivity of risk genes and a potential synaptic basis for schizophrenia.* Molecular Psychiatry, 2006: p. 1-13.
133. Kamiya, A., et al., *A schizophrenia-associated mutation of DISC1 perturbs cerebral cortex development.* Nature Cell Biology, 2005. **7**(12): p. 1067-78.
134. Schurov, I., Handford, EJ, Brandon, NJ, Whiting, PJ, *Expression of disrupted in schizophrenia 1 (DISC1) protein in the adult and developing mouse brain indicates its role in neurodevelopment.* Molecular Psychiatry, 2004. **9**: p. 1100-1110.
135. Duan X, Chang J., Ge S, Faulkner RL, Kim JY, Kitabatake Y, Liu XB, Yang CH, Jordan JD, Ma DK, Liu CY, Ganesan S, Cheng HJ, Ming GL, Lu B, Song H., *Disrupted-In-Schizophrenia 1 regulates integration of newly generated neurons in the adult brain.* Cell., 2007. **130**(6): p. 1146-1158.
136. Lipska, B., Peters, T, Hyde, TM, Halim, N, Horowitz, C, Mitkus, S, Weickert, CS, Matsumoto, M, Sawa, A, Straub, R, Vakkalanka, R, Herman, MM, Weinberger, DR, Kleinman, JE, *Expression of DISC1 binding partners is reduced in schizophrenia and associated with DISC1 SNPs.* Hum Mol Genet, 2006. **15**: p. 1245–1258.
137. Sasaki S, Shionoya A., Ishida M, Gambello MJ, Yingling J, and W.-B.A.e. al., *A LIS1/NUDEL/cytoplasmic dynein heavy chain complex in the developing and adult nervous system.* Neuron, 2000. **28**: p. 681-696.

138. Feng Y, Olson E., Stukenberg PT, Flanagan LA, Kirschner MW, and W. CA., *LIS1 Regulates CNS Lamination by Interacting with mNudE, a Central Component of the Centrosome.* Neuron, 2000. **28**: p. 665–679.
139. Feng Y, Walsh C., *Mitotic spindle regulation by Nde1 controls cerebral cortical size.* Neuron, 2004. **44**: p. 279–293.
140. Sasaki S, Mori D., Toyo-oka K, Chen A, Garrett-Beal L, Muramatsu M et al. , *Complete loss of Ndel1 results in neuronal migration defects and early embryonic lethality.* Mol Cell Biol, 2005. **25**: p. 7812-7827.
141. Hirotsune S, Fleck M., Gambello MJ, Bix GJ, Chen A, Clark GD et al. , *Graded reduction of Pafah1b1 (Lis1) activity results in neuronal migration defects and early embryonic lethality.* . Nature Genetics, 1998. **19**: p. 333-339.
142. Gambello MJ, Darling D., Yingling J, Tanaka T, Gleeson JG, and W.-B. A., *Multiple dose-dependent effects of Lis1 on cerebral cortical development.* . J Neurosci, 2003. **23**: p. 1719-1729.
143. Kamiya, A., Tomoda, T, Chang, J, Takaki, M, Zhan, C, Morita, M, Cascio, MB, Elashvili, S, Koizumi, H, Takanezawa, Y, Dickerson, F, Yolken, R, Arai, H, Sawa, A, *DISC1–NDEL1/NUDEL protein interaction, an essential component for neurite outgrowth, is modulated by genetic variations of DISC1.* Human Molecular Genetics, 2006. **15**(22): p. 3313-3323.
144. Lipska, B., Peters, T, Hyde, TM, Halim, N, Horowitz, C, Mitkus, S, Weickert, CS, Matsumoto, M, Sawa, A, Straub, R, Vakkalanka, R, Herman, MM, Weinberger, DR, Kleinman, JE, *Expression of DISC1 Binding Partners Is Reduced in Schizophrenia and Associated with DISC1 SNPs.* advanced publishing, 2006.
145. Hennah W, Tomppo L., Hiekkalinna T, Palo OM, Kilpinen H, Ekelund J et al. , *Families with the risk allele of DISC1 reveal a link between schizophrenia and another component of the same molecular pathway, NDE1.* . Hum Mol Genet, 2007. **16**: p. 453-462.
146. Tomppo L, Hennah W., Lahermo P, Loukola A, Ekelund J, Partonen T et al., *Association evidence from NUDEL and PDE4D support the DISC1-pathway concept in the etiology of schizophrenia.* American journal of medical genetics, 2006. **141B**.
147. Shinoda, T., Taya, S, Tsuboi, D, Hikita, T, Matsuzawa, R, Kuroda, S, Iwamatsu, A and Kaibuchi, K, *DISC1 Regulates Neurotrophin-Induced Axon Elongation via Interaction with Grb2.* The Journal of Neuroscience, 2007. **27**(1): p. 4-14.
148. Beard MB, O'Connell J., Bolger GB, Houslay MD. , *The unique N-terminal domain of the cAMP phosphodiesterase PDE4D4 allows for interaction with specific SH3 domains.* . FEBS Letters, 1999. **460**: p. 173–177.
149. Liu YF, Deth R., Devys D. , *SH3 domain-dependent association of huntingtin with epidermal growth factor receptor signaling complexes.* . J Biol Chem, 1997. **272**: p. 8121–8124.
150. Venezia V, Russo C., Repetto E, Nizzari M, Violani E, Carlo P et al., *Apoptotic cell death and amyloid precursor protein signaling in neuroblastoma SH-SY5Y cells.* Ann NY Acad Sci, 2004. **1030**: p. 339–347.

151. Nizzari M, Venezia V., Repetto E, Caorsi V, Magrassi R, Gagliani MC et al. , *Amyloid precursor protein and Presenilin1 interact with the adaptor GRB2 and modulate ERK 1,2 signaling.* J Biol Chem 2007, 2007. **282**: p. 13833–13844.
152. Oldenhof J, Vickery R., Anafi M, Oak J, Ray A, Schoots O et al., *SH3 binding domains in the dopamine D4 receptor.* Biochemistry, 1998. **37**: p. 15726–15736.
153. McShea A, Zelasko D., Gerst JL, Smith MA., *Signal transduction abnormalities in Alzheimer's disease: evidence of a pathogenic stimuli.* Brain Res., 1999. **815**(2): p. 237-42.
154. Dammermann A, Merdes A., *Assembly of centrosomal proteins and microtubule organization depends on PCM-1.* . J Cell Biol. , 2002 **159**(2): p. 255-266. .
155. Kamiya A, Tomoda P., Kubo K, Engelhard C, Ishizuka K, Kubo A, Tsukita S, Pulver AE, Nakajima K, Cascella NG, Katsanis N, Sawa A., *Recruitment of PCM1 to the centrosome by the cooperative action of DISC1 and BBS4: a candidate for psychiatric illnesses.* . Arch Gen Psychiatry., 2008 **65**(9): p. 996-1006.
156. Gurling HMD, Critchley H., Datta SR, McQuillin A, Blaveri E, Thirumalai S, Pimm J, Krasucki R, Kalsi G, Quested D, Lawrence J, Bass N, Choudhury K, Puri V, O'Daly O, David Curtis D, Blackwood D, Muir W, Malhotra AK, Buchanan RW, Good CD, Frackowiak RSJ, Dolan RJ. , *Genetic Association and Brain Morphology Studies and the Chromosome 8p22 Pericentriolar Material 1 (PCM1) Gene in Susceptibility to Schizophrenia* Arch Gen Psychiatry 2006 **63**(8): p. 844-854.
157. Zhu L, Li D., Hu J, Cheng J, Wang S, Wang Q, Wang F, Chen J, Wang J *GSK-3β Inhibits Presynaptic Vesicle Exocytosis by Phosphorylating P/Q-Type Calcium Channel and Interrupting SNARE Complex Formation* The Journal of Neuroscience, 2010. **30**(10): p. 3624-3633.
158. Mao Y, Ge X., Frank CL, Madison JM, Koehler AN, Doud MK, Tassa C, Berry EM, Soda T, Singh KK, Biechele T, Petryshen TL, Moon RT, Haggarty SJ, Tsai LH., *Disrupted in schizophrenia 1 regulates neuronal progenitor proliferation via modulation of GSK3beta/beta-catenin signaling.* Cell., 2009. **136**(6): p. 1017-1031.
159. Arguello PA, Gogos J., *A signalling pathway AKTing up in schizophrenia.* J Clin Invest, 2008. **118**(6): p. 2018-2021.
160. Kim JY, Duan X., Liu CY, Jan MH, Guo JU, Pow-anpongkul N, Kang E, Song H, Ming GL *DISC1 regulates new neuron development in the adult brain via modulation of AKT-mTOR signaling through KIAA1212.* Neuron 2009. **63**: p. 761-773.
161. Enomoto A, Murakami H, Asai N, Morone N, Watanabe T, Kawai K, Murakumo Y, Usukura J, Kaibuchi K, Takahashi M., *Akt/PKB regulates actin organization and cell motility via Girdin/APE.* . Dev Cell. , 2005 **9**(3): p. 389-402.
162. Honda A, Miyoshi K., Baba K, Taniguchi M, Koyama Y, Kuroda S et al. , *Expression of fasciculation and elongation protein zeta-1 (FEZ1) in the developing rat brain.* Brain Res Mol Brain Res, 2004. **122**: p. 89–92.
163. K Miyoshi, Honda A, K Baba, M Taniguchi, K Oono, T Fujita, Kuroda, T Katayama, and M Tohyama, *Disrupted-In-Schizophrenia 1, a candidate gene for*

schizophrenia, participates in neurite outgrowth. Molecular Psychiatry, 2003. **8**: p. 685–694.
164. Koga M, Ishiguru H., Horiuchi Y, Albalushi T, Inada T, Iwata N et al., *Failure to confirm the association between the FEZ1 gene and schizophrenia in a Japanese population.* Neurosci Letters, 2007. **417**: p. 326–329.
165. Hodgkinson CA, Goldman D., Ducci F, DeRosse P, Caycedo DA, Newman ER et al., *The FEZ1 gene shows no association to schizophrenia in Caucasian or African American populations.* Neuropsychopharmacology, 2007. **32**: p. 190–196.
166. Yamada K, Nakamura K., Minabe Y, Iwayama-Shigeno Y, Takao H, Toyota T et al. , *Association analysis of FEZ1 variants with schizophrenia in Japanese cohorts.* . Biol Psychiatry, 2004. **56**: p. 683–690.
167. Morris JA, Kandpal G., Ma L, Austin CP, *DISC1 (Disrupted-In-Schizophrenia 1) is a centrosome-associated protein that interacts with MAP1A, MIPT3, ATF4/5 and NUDEL: regulation and loss of interaction with mutation.* . Hum Mol Genet, 2003. **12**: p. 1591–1608.
168. Sawamura N, Ando T., Maruyama Y, Fujimuro M, Mochizuki H, Honjo K, Shimoda M, Toda H, Sawamura-Yamamoto T, Makuch LA, Hayashi A, Ishizuka K, Cascella NG, Kamiya A, Ishida N, Tomoda T, Hai T, Furukubo-Tokunaga K,Sawa A *Nuclear DISC1 regulates CRE-mediated gene transcription and sleep homeostasis in the fruit flyDISC1 in transgenic flies and its nuclear function.* Molecular Psychiatry, 2008. **13** p. 1138-1148.
169. Talukder AH, Vadlamundi R., Mandal M, Kumar R., *Heregulin induces expression, DNA binding activity, and transactivating functions of basic leucine zipper activating transcription factor 4.* Cancer Res, 2000. **60**: p. 276–281.
170. Zhong Y, Wu C., *Altered synaptic plasticity in Drosophilia memory mutants with a defective cyclic AMP cascade.* Science, 1991. **251**(4990): p. 198-201.
171. Chen CN, Denome S., Davis RL., *Molecular analysis of cDNA clones and the corresponding genomic coding sequences of the Drosophila dunce+ gene, the structural gene for cAMP phosphodiesterase.* Proc Nat Acad Sci USA, 1986. **83**(24): p. 9313-9317.
172. James R, Adams R., Christie S, Buchanan SR, Porteous DJ, Millar JK., *Disrupted in Schizophrenia 1 (DISC1) is a multicompartmentalized protein that predominantly localizes to mitochondria.* Molecular & Cellular Neurosciences, 2004. **26**(1): p. 112-122.
173. Sawamura N, Sawamura-Yamoto T., Ozeki Y, Ross CA, Sawa A., *A form of DISC1 enriched in nucleus: altered subcellular distribution in orbitofrontal cortex in psychosis and substance/alcohol abuse.* Proc Nat Acad Sci USA, 2005. **102**(4): p. 1187-1192.
174. Trichard C, Paillere-Martinot M., Attar-Levy D, Blin J, Feline A, Martinot JL, .*No serotonin 5-HT2A receptor density abnormality in the cortex of schizophrenic patients studied with PET.* . Schizophr Res. , 1998. **31**(1): p. 13-17.
175. Arranz MJ, Munro J., Sham P, Kirov G, Murray RM, Collier DA, Kerwin RW., *Meta-analysis of studies on genetic variation in 5-HT2A receptors and clozapine response.* . Schizophr Res., 1998 **32:**(2): p. 93-99.
176. Javitt DC, Zukin S., *Recent advances in the phencyclidine model of schizophrenia.* . Am J Psychiatry.;, 1991. **148**(10): p. 1301-1308.

177. Lahti AC, Koffel B., LaPorte D, Tamminga CA. , *Subanesthetic doses of ketamine stimulate psychosis in schizophrenia.* . Neuropsychopharmacology., 1995 **13**(1): p. 9-19.
178. Malhotra AK, Pinals D., Adler CM, Elman I, Clifton A, Pickar D, Breier A., *Ketamine-induced exacerbation of psychotic symptoms and cognitive impairment in neuroleptic-free schizophrenics.* Neuropsychopharmacology. , 1997. **17**(3): p. 141-150.
179. Breier A, Adler C., Weisenfeld N, Su TP, Elman I, Picken L, Malhotra AK, Pickar D., *Effects of NMDA antagonism on striatal dopamine release in healthy subjects: application of a novel PET approach.* . Synapse. , 1998 **29**(2): p. 142-147.
180. Kornhuber J, Mack-Burkhardt F., Riederer P, Hebenstreit GF, Reynolds GP, Andrews HB, Beckmann H. , *[3H]MK-801 binding sites in postmortem brain regions of schizophrenic patients.* . J Neural Transm. , 1989;. **77**(2-3): p. 231-236.
181. Sherman AD, Davidson A., Baruah S, Hegwood TS, Waziri R., *Evidence of glutamatergic deficiency in schizophrenia.* . Neurosci Lett. , 1991 **121**(2): p. 77-80.
182. Li D, He L., *Association study between the NMDA receptor 2B subunit gene (GRIN2B) and schizophrenia; a HuGE review and meta-analysis.* Genet Med, 2007. **9**(1): p. 4-8.
183. van Rossum, J., *The significance of dopamine-receptor blockade for the mechanism of action of neuroleptic drugs.* Arch In Pharmacodyn, 1966. **160**(2): p. 492-494.
184. Abi-Dargham A, Rodenhiser J., Printz D, Zea-Ponce Y, Gil R, Kegeles LS, Weiss R, Cooper TB, Mann JJ, Van Heertum RL, Gorman JM, Laruelle M., *Increased baseline occupancy of D2 receptors by dopamine in schizophrenia.* . Proc Natl Acad Sci U S A. , 2000 **97**(14): p. 8104-8109.
185. Seeman P, Kapur S., *Schizophrenia: More dopamine, more D2 receptors* Proc Natl Acad Sci U S A. , 2000 **97**(14): p. 7673-7675.
186. Kellendonk C, Simpson E., Jonathan Polan HJ, Malleret GI, Vronskaya S, Winiger V, Moore H, Kandel ER, *Transient and Selective Overexpression of Dopamine D2 Receptors in the Striatum Causes Persistent Abnormalities in Prefrontal Cortex Functioning.* Neuron, 2006. **49**: p. 603-615.
187. Glahn DC, Therman S., Manninen M, Huttunen M, Kaprio J, Lönnqvist J, Cannon TD., *Spatial working memory as an endophenotype for schizophrenia.* Biol Psychiatry. , 2003. **53**(7): p. 624-626.
188. Snyder PJ, Jackson C., Piskulic D, Olver J, Norman T, Maruff P., *Spatial working memory and problem solving in schizophrenia: the effect of symptom stabilization with atypical antipsychotic medication.* Psychiatry Res. , 2008. **160**(3): p. 316-326.
189. Weinberger DR, *The Pathogenesis of schizophrenia: a neurodevelopmental theory.* The neurobiology of schizophrenia, 1986: p. 387-405.
190. Niemi LT, Suvisaari J., Tuulio-Henriksson A, Lonnqvist JK, *Childhood developmental abnormalities in schizohrenia: evidence from high-risk studies.* Schizophr Res., 2003. **60**(2-3): p. 239-258.

191. Keshavan MS, Diwadker V., Montrose DM, Rajarethinam R, Sweeney JA, *Premorbid indicators and risk for schizophrenia: a selective review and update.* Schizophrenia research, 2005. **79**: p. 45-57.
192. James ACD, Crow T., Renowden S, Wardell AMJ, Smith DM, Anslow R *Is the course of brain development in schizophrenia delayed? Evidence from onsets in adolescence.* Schizophrenia Research 1999 **40**(1): p. 1-10.
193. Hikida T, Jaaro-Peled H., Seshadri S, Oishi K, Hookway C, Kong S, Wu D, Xue R, Andradé M, Tankou S, Mori S, Gallagher M, Ishizuka K, Pletnikov M, Kida S, Sawa A, *Dominant-negative DISC1 transgenic mice display schizophrenia-associated phenotypes detected by measures translatable to humans.* PNAS, 2007. **104**(36): p. 14501-14506.
194. Niwa M, Kamiya A., Murai R, Kubo K, Gruber AJ, Tomita K, Lu L, Tomisato S, Jaaro-Peled H, Seshadri S, Hiyama H, Huang B, Kohda K, Noda Y, O'Donnell P, Nakajima K, Sawa A, Toshitaka Nabeshima T, *Knockdown of DISC1 by In Utero Gene Transfer Disturbs Postnatal Dopaminergic Maturation in the Frontal Cortex and Leads to Adult Behavioral Deficits.* Neuron, 2010. **65**: p. 480-489.
195. Davis RL, Dauwalder B., *The Drosophila dunce locus: learning and memory genes in the fly* Trends in Genetics, 1991. **7**(7): p. 224-229.
196. Siuciak JA, Chapin D., McCarthy SA, Martin AN, *Antipsychotic profile of rolipram: efficacy in rats and reduced sensitivity in mice deficient in the phosphodiesterase-4B (PDE4B) enzyme.* Psychopharmacology, 2007. **192**(3): p. 415-424.
197. Clapcote, S., Lipina, TV, Millar, JK, Mackie, S, Christie, S, Ogawa, F, Lerch, JP, Trimble, K, Uchiyama, M, Sakuraba, Y, Kaneda, H, Shiroishi, T, Houslay, MD, Henkelman, RM, Sled, JG, Gondo, Y, Porteous, DJ, Roder, JC, *Behavioral Phenotypes of Disc1 Missense Mutations in Mice.* Neuron, 2007. **54**: p. 387–402.
198. Stranger BE, Forrest M., Dunning M, Ingle CE, Beazley C, et al. , *Relative impact of nucleotide and copy number variation on gene expression phenotypes. .* Science 2007. **315**: p. 848–853.
199. Torkamani A, Dean B., Schork NJ, Thomas EA, *Coexpression network analysis of neural tissue reveals perturbations in developmental processes in schizophrenia.* Genome Research, 2010. **20**: p. 403-412.
200. Harris LW, Lockstone H., Khaitovich P, Weickert CS, Webster MJ, Bahn S, *Gene expression in the prefrontal cortex during adolescence: Implications for the onset of schizophrenia.* BMC Medical Genomics, 2009. **2**(28).
201. Tordjman S, Drapier D., Bonnot O, Graignic R, Fortes S, Cohen D, Millet B, Laurent C, Roubertoux PL., *Animal models relevant to schizophrenia and autism: validity and limitations.* Behavioural Genetics, 2007. **37**(1): p. 61-70.
202. Cannon TD, Keller M., *Endophenotypes in the genetic analysis of mental disorders.* Annual Reviews Clinical Psychology, 2006. **2**: p. 267-290.
203. Gould TD, Gottessman I., *Psychiatric endophenotypes and the development of valid animal models.* Genes, Brain and Behaviour, 2006. **5**(2): p. 113-119.
204. Gill M, Donohoe G and Corvin A., *What have the genomics ever done for the psychoses?* Psychological Medicine, 2010. **40**: p. 529–540.

205. Sebat J, Levy DL, McCarthy SE, *Rare structural variants in schizophrenia:one disorder, multiple mutations; one mutation, multiple disorders.* Trends in Genetics, 2010. **25**(12): p. 528-535.
206. Koike, H., et al., *Disc1 is mutated in the 129S6/SvEv strain and modulates working memory in mice.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(10): p. 3693-3697.
207. Kvajo M, McKeller H., Arguello PA, Drew LJ, Moore H, MacDermott AB, Karayiorgou M, Gogos JA., *A mutation in mouse Disc1 that models a schizophrenia risk allele leads to specific alterations in neuronal architecture and cognition.* Proc Natl Acad Sci U S A., 2008. **105**(19): p. 7076-7081.
208. Goldman-Rakic, P., *Working memory dysfunction in schizophrenia.* J Neuropsychiatry Clin Neurosci, 1994. **6**: p. 348-357.
209. Arnold SE, Ruscheinsky D., Han L-Y, *Further evidence of abnormal crtoarchitecture of the entorhinal cortex in schizophrenia using spacial point pattern analysis.* Biological psychiatry, 1997. **42**: p. 639-647.
210. Clapcote , S., Roder, JC, *Deletion Polymorphism of Disc1 is Common to All 129 Mouse Substrains: Implications for Gene Targeting Studies of Brain Function.* Genetics, 2006.
211. Lawrie SM, Abukmeil S., *Brain abnormality in schizophrenia. A systematic and quantitative review of volumetric magnetic resonance imaging studies* The British Journal of Psychiatry, 1998. **172**: p. 110-120.
212. Pletnikov MV, Ayhan Y., Xu Y, Nikolskaia O, Ovanesov M, Huang H, Mori S, Moran TH, Ross CA, *Enlargement of the lateral ventricles in mutant DISC1 transgenic mice.* Molecular Psychiatry, 2008. **13**(115): p. 173–186.
213. Ayhan Y, Abazyan B., Nomura J, Kim R, Ladenheim B, Krasnova IN, Sawa A, Margolis RL, Cadet JL, Mori S, Vogel MW, Ross CA, Pleitnikov MW, *Differential effects of prenatal and postnatal expressions of mutant human DISC1 on neurobehavioural phenotypes in transgenic mice: evidence for neurodevelopmental origin of major psychiatric disorders.* Mol Psychiatry, 2010. **Epub ahead of print**.
214. Shen S, Lang B., Nakamoto C, Zhang F, Pu J, Kuan S, Chatzi C, He S, Mackie I, Brandon NJ, Marquis KL, Day M, Hurko O, McCaig CD, Riedel G, St Clair D, *Schizophrenia-Related Neural and Behavioral Phenotypes in Transgenic Mice Expressing Truncated Disc1*. The Journal of Neuroscience, 2008. **28**(43): p. 10893-10904.
215. Kuhn K, Baker S., Chudin E, Lieu MH, Oeser S, Bennett H, Rigault P, Barker D, McDaniel TK, Chee MS., *A novel, high-performance random array platform for quantitative gene expression profiling.* Genome Res, 2004. **14**(11): p. 2347-2356.
216. Du P, Kibbe W., Lin SM, *lumi: a pipeline for processing Illumina microarray.* Bioinformatics, 2008. **24**(13): p. 1547-1458.
217. Kirov G, Grozeva D., Norton N, Ivanov D, Mantripragada KK, Holmans P; International Schizophrenia Consortium; Wellcome Trust Case Control Consortium, Craddock N, Owen MJ, O'Donovan MC., *Support for the involvement of large copy number variants in the pathogenesis of schizophrenia.* Human Molecular Genetics, 2009. **18**(8): p. 1497-1503.

218. Kumari V, Antonova E., Geyer MA, Ffytche D, Williams SC, Sharma T, *A fMRI investigation of startle gating deficits in schizophrenia patients treated with typical or atypical antipsychotics.* Int J Neuropsychopharmacol, 2007. **4**(10): p. 463-477.
219. Swerdlow NR, Geyer M., *Using an animal model of deficient sensorimotor gating to study the pathophysiology and new treatments of schizophrenia.* Schizophrenia Bulletin, 1998. **24**(2): p. 285-301.
220. Braff DL, Geyer M., Swerdlow NR, *Human studies of PrePulse Inhibition of startle: patient groups and pharmacological studies.* Psychopharmacology, 2001(156): p. 234-258.
221. Swerdlow NR, Geyer M., Hartman PL, Sprock J, Auerbach PP, Cadenhead K, Perry W, Braff DL, *Sex differences in sensorimotor gating of the human startle reflex: all smoke?* Psychopharmacology, 1999(146): p. 228-232.
222. Swerdlow NR, Geyer M., Braff DL, *Neural Circuitry of PrePulse Inhibition of Startle in the Rat: Current Knowledge and Future Challenges.* Psychopharmacology, 2001(156): p. 194 - 215.
223. Matsuki, T, Hori G, Furuichi T, *Gene expression profiling during the embryonic development of mouse brain using an oligonucleotide-based microarray system.* Molecular Brain Research, 2005. **136**: p. 231-254.
224. Qui, W, Lee MT, et al, *Sample size and power calculation in microarray studies using the sizepower package.* R Package Version 1.6.0, 2007.
225. Shi L, Reid LH, et al, *The microarray quality control (MAQC) project shows inter and intra-platform reproducibility of gene expression measurements.* Nat Biotechnol, 2006. **24**(9): p. 1151-1161.
226. Schroeder A, Mueller O.et al, *The RIN: an RNA integrity number for assigning integrity values to RNA measurements.* BMC Mol Biol, 2006. **7**: p. 3.
227. Imbeaud S, Graudens E, et al, *Towards standardisation of RNA quality assessment using user-dependent classifiers of microcapillary electrophoresis traces.* Nucleic Acid Res, 2005. **33**(6): p. 395-399.
228. Jones L, Goldstein D., Hughes G, Strand AD, Collin F, Dunnett SB, Kooperberg C, Aragaki A, Olson JM, Augood SJ, Faull RL, Luthi-Carter R, Moskvina V, Hodges AK, *Assessment of the relationship between pre-chip and post-chip quality measures for Affymetrix GeneChip expression data.* BMC Bioinformatics, 2006. **7**(211).
229. Dunning MJ, Thorne NP, *Quality control and low-level statistical analysis of illumina bead arrays.* REVSTAT, 2006. **4**(1): p. 1-30.
230. Quakenbush, J. *Computational analysis of microarray data.* Nat Rev Genet, 2001. **2**(6): p. 418-427.
231. Su AI, Cooke MP, et al, *Large-scale analysis of the human and mouse transcriptomes.* Proc Nat Acad Sci USA, 2002. **99**(7): p. 4465-4470.
232. Paylor R, Crawley J., *Inbred strain differences in perpulse inhibition of the mouse startle response.* . Psychopharmacology, 1997. **132**: p. 169-180.
233. Hall J, Whalley H., Marwick K, McKirdy J, Sussmann J, Romaniuk L, Johnstone EC, Wan HI, McIntosh AM, Lawrie SM, *Hippocampal function in schizophrenia and bipolar disorder.* Psychological medicine, 2010. **40**(5): p. 761-770.

234. Sigurdsson T, Stark K., Karayiorgou M, Gogos JA, Gordon JA, *Impaired hippocampal–prefrontal synchrony in a genetic mouse model of schizophrenia.* Nature, 2010. **464**: p. 763-767.
235. Wu H, Yang H., Churchill GA, *R/MAANOVA: An extensive R environment for the analysis of microarray experiments.* 2008.
236. Cui X, Hwang J., Qui J, Bades NJ, Churchill GA, *Improved statistical tests for differential gene expression by shrinking variance components estimates.* Biostatistics, 2005. **6**(1): p. 59-75.
237. Christoforou A, LeHellard S., Thomson PA, Morris SW, Tenesa A, Pickard BS, Wray NR, Muir WJ, Blackwood DH, Porteous DJ and Evans KL, *Association analysis of the chromosome 4p15–p16 candidate region for bipolar disorder and schizophrenia.* Molecular Psychiatry, 2008. **12**: p. 1011–1025.
238. Yamada K, Gerber D., Iwayama Y, Ohnishi T, Ohba H, Toyota H, Aruga J, Minabe Y, Tonegawa S, Yoshikawa T *Genetic analysis of the calineurin pathway identifies members of the EGR gene family, specifically EGR3, as potential susceptibility candidates in schizophrenia.* PNAS, 2006. **104**: p. 2815-2820.
239. Ching MS, Shen Y., Tan WH, Jeste SS, Morrow EM, Chen X, Mukaddes NM, Yoo SY, Hanson E, Hundley R, Austin C, Becker RE, Berry GT, Driscoll K, Engle EC, Friedman S, Gusella JF, Hisama FM, Irons MB, Lafiosca T, LeClair E, Miller DT, Neessen M, Picker JD, Rappaport L, Rooney CM, Sarco DP, Stoler JM, Walsh CA, Wolff RR, Zhang T, Nasir RH, Wu BL;, *Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders.* Am J Med Genet B Neuropsychiatr Genet., 2010 **153B**(4): p. 937-947.
240. Sullivan, P., Fan, C, Perou, CM, *Evaluating the comparability of gene expression in blood and brain.* American journal of medical genetics part B (neuropsychiatric genetics), 2006(141B): p. 261-268.
241. Larsson O, Wahlestedt C., Timmons JA, *Considerations when using the significance analysis of microarrays (SAM) algorithm.* BMC Bioinformatics, 2005. **29**(6).
242. Effron T and Tibshirani R, *Empirical bayes methods and false discovery rates for microarrays.* Genet. Epidemiol., 2002. **23**: p. 70-86.
243. Curtis RK, Oresic M and Vidal-Puig A., *Pathways to the analysis of microarray data.* Trends in biotechnology, 2005. **23**(8): p. 429-435.
244. Bertsch B, Ogden CA, Sidhu K, Le-Niculescu H, Kuczenski R and Niculescu AB, *Convergent Functional Genomics: A Baysian candidate gene identification approach for complex disorders.* Science, 2005. **Methods 37**: p. 274-279.
245. VanGuilder HD, Vrana K., Freeman WM., *Twenty-five years of quantitative PCR for gene expression analysis.* Biotechniques, 2008. **44**(5): p. 619-626.
246. Canales RD, Luo Y., Willey JC, Austermiller B, Barbacioru CC, Boysen C, Hunkapiller K, Jensen RV, Knight CR, Lee KY, Ma Y, Maqsodi B, Papallo A, Peters EH, Poulter K, Ruppel PL, Samaha RR, Shi L, Yang W, Zhang L, Goodsaid FM., *Evaluation of DNA microarray results with quantitative gene expression platforms.* National Biotechnology, 2006. **24**(9): p. 1115-1122.
247. Valasek MA, Repa J., *The power of real-time PCR.* Advan Physiol Educ 2005. **29**: p. 151-159.

248. Simon R, Radmacher M., Dobbin K, McShane LM., *Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification.* J Natl Cancer Inst., 2003. **95**(1): p. 14-18.
249. Sudhof, T., *Neuroligins and neurexins link synaptic function to cognitive disease.* Nature, 2008. **455**: p. 903-911.
250. Craig, A., and Kang, Y, *Neurexin–neuroligin signaling in synapse development.* Current Opinion in Neurobiology, 2007. **17**: p. 43-52.
251. Need AC, Ge D, Weale ME, Maia J, Feng S, Heinzen E, et al., *A Genome-Wide Investigation of SNPs and CNVs in Schizophrenia.* PLoS Genetics, 2009. **5**(2): p. e1000373. doi:10.1371/journal.pgen.1000373.
252. Vrijenhoek T, Buizer-Voskamp J., van der Stelt I, Strengman E; Genetic Risk and Outcome in Psychosis (GROUP) Consortium, Sabatti C, Geurts van Kessel A, Brunner HG, Ophoff RA, Veltman JA., *Recurrent CNVs disrupt three candidate genes in schizophrenia patients.* American Journal of Human Genetics, 2008. **83**(4): p. 504-510.
253. Rujescu D, Ingason A., Cichon S, Pietiläinen OP, Barnes MR, Toulopoulou T, Picchioni M, Vassos E, Ettinger U, Bramon E, Murray R, Ruggeri M, Tosato S, Bonetto C, Steinberg S, Sigurdsson E, Sigmundsson T, Petursson H, Gylfason A, Olason PI, Hardarsson G, Jonsdottir GA, Gustafsson O, Fossdal R, Giegling I, Möller HJ, Hartmann AM, Hoffmann P, Crombie C, Fraser G, Walker N, Lonnqvist J, Suvisaari J, Tuulio-Henriksson A, Djurovic S, Melle I, Andreassen OA, Hansen T, Werge T, Kiemeney LA, Franke B, Veltman J, Buizer-Voskamp JE; GROUP Investigators, Sabatti C, Ophoff RA, Rietschel M, Nöthen MM, Stefansson K, Peltonen L, St Clair D, Stefansson H, Collier DA., *Disruption of the neurexin 1 gene is associated with schizophrenia.* Human Molecular Genetics, 2009. **18**(5): p. 988-996.
254. Feng J, Schroer R., Yan J, Song W, Yang C, Bockholt A, Cook EH Jr, Skinner C, Schwartz CE and Sommer SS *High frequency of neurexin 1 beta signal peptide structural variants in patients with autism.* Neuroscience Letters, 2006. **409**: p. 10-13.
255. Novak G, Boukhadra J., Shaikh SA, Kennedy JL, Le Foll B., *Association of a polymorphism in the NRXN3 gene with the degree of smoking in schizophrenia: a preliminary study.* World J Biol Psychiatry, 2009. **10**(4/3): p. 929-935.
256. Hishimoto A, Lui Q., Drgon T, Pletnikova O, Walther D, Zhu X, Troncoso JC and Uhl GR *Neurexin 3 polymorphisms are associated with alcohol dependence and altered expression of specific isoforms.* Human Molecular Genetics, 2007. **16**(23): p. 2880-2891.
257. Cheng MC, Liao D., Hsiung CA, Chen CY, Liao YC, Chen CH. , *Chronic treatment with aripiprazole induces differential gene expression in the rat frontal cortex.* Int J Neuropsychopharmacol, 2007: p. 1-10.
258. Mulligan MK, Ponomarev I., Boehm SL, Owen JA, Levin PS, Berman AE, Blednov YA, Crabbe JC, Williams RW, Miles MF and Bergeson SE *Alcohol trait and transcriptional genomic analysis of C57BL/6 substrains.* . Genes, Brain and Behaviour, 2008. **7**: p. 677-689.
259. Blatner NR, Stahelin R., Diraviyam K, Hawkins PT, Hong W, Murray D, and Cho W, *The Molecular Basis of the Differential Subcellular Localization of FYVE*

Domains. THE JOURNAL OF BIOLOGICAL CHEMISTRY, 2004. **279**(51): p. 53818-53827

260. Bienvenu T, des Portes V., McDonnell N, Carrie A, Zemni R, Couvert P, Ropers HH, Moriane C, van Bokhoven H, Fryns JP, Allen K, Walsh CA, Boue J, Kahn A, Chelly J, Beldjord C, *Missense mutantion in PAK3, R67C, causes X-linked non-specific mental retardation.* American Journal of Medical Genetics, 2000. **93**: p. 294-298.
261. Donnelly AJ, Partington M., Ryan AK, Mulley JC, *Regional localisation of two non-specific X-linked mental retardation genes (MRX30 and MRX31).* American Journal of Medical Genetics, 1996. **64**: p. 113-120.
262. Peippo M, Koivisto A., Sarkamo T, Sipponen M, von Koskull H, Ylisaukko-oja T, Rehnstrom K, Froyen G, Ignatius J and Jarvela I *PAK3 Related Mental Disability: Further Characterization of the Phenotype.* American Journal of Medical Genetics Part A, 2007. **143A**: p. 2406-2416.
263. Boda B, Alberi S., Nikonenko I, Node-Langois R, Jourdain P, Moosmayer M, Parisi-Jourdin L and Muller D *The Mental Retardation Protein PAK3 Contributes to Synapse Formation and Plasticity in Hippocampus.* The Journal of Neuroscience, 2004. **24**(48): p. 10816-10825
264. Meng J, Meng Y., Hanna A, Janus C and Jia Z *Abnormal Long-Lasting Synaptic Plasticity and Cognition in Mice Lacjking the Mental Retardatio Gene PAK3.* The Journal of Neuroscience, 2005. **25**(28): p. 6641-6650.
265. Penzes P, Jones K., *Dendritic spine dynamics--a key role for kalirin-7.* Trends in Neurosciences, 2008. **31**(8): p. 419-427.
266. Suzuki S, Sano K and Taniharat H., *Diversity of the cadherin family: evidence for eight new cadherins in nervous tissue.* Cell Regulation, 1991. **2**: p. 261-270.
267. AKINS MR, BENSON D., AND GREER CA, *Cadherin Expression in the Developing Mouse Olfactory System.* THE JOURNAL OF COMPARATIVE NEUROLOGY, 2007. **501**: p. 483-497.
268. Choi P, Jordan C., Mendez E, Houck J, Yueh B, Farwell DG, Futran N, Chen C, *Examination of oral cancer biomarkers by tissue microarray analysis.* Archives of Otolaryngology - head and neck surgery, 2008. **134**(5): p. 539-546.
269. Nakajima G, Patino-Garcia A., Bruheim S, Xi Y, San Julian M, Lecanda F, Sierrasesumaga L, Müller C, Fodstad O, Ju J., *CDH11 expression is associated with survival in patients with osteosarcoma.* Cancer Genomics and Proteomics, 2008. **5**(1): p. 37-42.
270. Binder EB, Kinkead B., Owens MJ, Kilts CD and Nemeroff CB, *Enhanced neurotensin neurotransmission is involved in the clinically relevant behavioural effects of antipsychotoc drugs: evidence from animal models of sensorimotor gating.* Journal of Neuroscience, 2001. **21**(2): p. 601-608.
271. Wilson HL, Wong A., Shaw SR, Tse W-Y, Stapleton GA, Phelan MC, Hu S, Marshall J, McDermid HE, *Molecular characterisation of the 22q13 deletion syndrome supports the role of haploinsufficiency of SHANK3/PROSAP2 in the major neurological symptoms.* Journal of Medical Genetics, 2003. **40**: p. 575–584.
272. Durand CM, Betancur C., Boeckers TM, Bockmann J, Chaste P, Fauchereau F, Nygren G, Rastam M, Gillberg C, Anckarsa¨ter H, Sponheim E, Goubran-Botros

H, Delorme R, Chabane N, Mouren-Simeoni MC, de Mas P, Bieth E, Roge B, He´ron D, Burglen L, Gillberg C, Leboyer M and Bourgeron T, *Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders.* Nature Genetics, 2007. **39**(1): p. 25-26.
273. Rifkin L, Lewis S., Jones P, Toone B, Murray R., *Low birth weight and schizophrenia.* Br J Psychiatry. , 1994 **165**(3): p. 357-362.
274. Brumwell CL, Curan T., *Developmental mouse brain gene expression maps.* Journal of Physiology, 2006. **575**(part2): p. 343-346.
275. Osheroff H, Hatten M., *Gene expression profiling of preplate neurons destined for the subplate: genes involved in transcription, axon extension, neurotransmitter regulation, steroid hormone signalling and neuronal survival.* Cerebral Cortex, 2009. **Suppl1**: p. 126-134.
276. Kirkpatrick B, Xu L., Cascella N, Ozeki Y, Sawa A, Roberts RC, *DISC1 immunoreactivity at the light and ultrastructural level in the human neocortex* THE JOURNAL OF COMPARATIVE NEUROLOGY, 2006. **497**(3): p. 436-450.
277. Chuaqui RF, Bonner R., Best CJM, Gillespie JW, Flaig MJ, Hewitt SM, Phillips JL, Krizman DB, Tangrea MA, Ahram M, Linehan WM, Knezevic V, Emmert-Buck MR, *Post-analysis follow-up and validation of microarray experiments.* nature genetics supplement, 2002. **32**: p. 509-514.
278. Gassmann M, Grenacher B., Rohde B, Vogel J, *Quantifying Western blots: Pitfalls of densitometry.* Electrophoresis, 2009. **30**: p. 1845-1855.
279. Beveridge, N.J et al, *Dysregulation of miRNA 181b in the temporal cortex in schizophrenia.* . Hum Mol Genet 2008. **17**: p. 1156-1168.
280. Lee HJ, Song J., Kim JW, Jin S, Suk Hong M, Park JK, Chung J, Shibata H, Fukumaki Y *Association study of polymorphisms in synaptic vesicle-associated genes, SYN2 and CPLX2, with schizophrenia* Behavioral and Brain Functions 2005. **1**(15).
281. Ching MS, Shen Y., Tan WH, Jeste SS, Morrow EM, Chen X, Mukaddes NM, Yoo SY, Hanson E, Hundley R, Austin C, Becker RE, Berry GT, Driscoll K, Engle EC, Friedman S, Gusella JF, Hisama FM, Irons MB, Lafiosca T, LeClair E, Miller DT, Neessen M, Picker JD, Rappaport L, Rooney CM, Sarco DP, Stoler JM, Walsh CA, Wolff RR, Zhang T, Nasir RH, Wu BL;,, *Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders.* . Am J Med Genet B Neuropsychiatr Genet., , 2010 **153B**(4): p. 937-947.
282. Morey JS, James C, Ryan J, Van Dolah FM *Microarray validation: factors influencing correlation between oligonucleotide microarrays and real-time PCR.* Biol Proced Online., 2006. **8**: p. 175–193.
283. Rajeevan MS, Vernon S., Taysavang N, Unger ER., *Validation of array-based gene expression profiles by real-time (kinetic) RT-PCR.* . J Mol Diagn. , 2001 **3**(1): p. 26-31.
284. Kuwabara, T. et al, *Wnt-mediated activation of NeuroD1 and retro-elements during adult neurogenesis.* . Nature Neurosci., 2009. **12**: p. 1097–1105
285. Gordon JA, *Testing the glutamate hypothesis of schizophrenia* Nature Neuroscience 2010 **13**: p. 2-4.
286. Hayashi-Takagi A, Takaki M., Graziane N, Seshadri S, Murdoch H, Dunlop AJ, Makino Y, Seshadri AJ, Ishizuka K, Srivastava DP, Xie Z, Baraban JM, Houslay

MD, Tomoda T, Brandon NJ, Kamiya A, Yan Z, Penzes P, Sawa A., *Disrupted-in-Schizophrenia 1 (DISC1) regulates spines of the glutamate synapse via Rac1*. Nat Neurosci., 2010. **13**(3): p. 327-332.
287. CLANCY B, DARLINGTON R., FINLAY BL *TRANSLATING DEVELOPMENTAL TIME ACROSS MAMMALIAN SPECIES* Neuroscience, 2001 **105** (1): p. 7-17.
288. Fairless R, Masius H., Rohlmann A, Heupel K, Ahmad M, Reissner C, Dresbach T, Missler M *Polarized Targeting of Neurexins to Synapses Is Regulated by their C-Terminal Sequences* The Journal of Neuroscience, 2008. **28**(48): p. 12969-1298.
289. Wang Y, Barbacioru C., Hyland F, Xiao W, Hunkapiller KL, Blake J, Chan F, Gonzalez C, Zhang L and Samaha R, *Large scale real-time PCR validation on gene expression measurements from two commercial long-oligonucleotide microarrays.* BMC Genomics 2006. **7**(59).
290. MacKenzie, S. and Housley MD, *Action of rolipram on specific PDE4 cAMP phosphodiesterase isoforms and on the phosphorylation of cAMP-response-element-binding protein (CREB) and p38 mitogen-activated protein (MAP) kinase in U937 monocytic cells.* Biochemistry, 2000. **347**(2): p. 571-578.
291. Jones, HM and Pilowsky LS, *Dopamine and antipsychotic drug action revisited.* The British Journal of Psychiatry, 2002. **181**: p. 271-275.
292. Duncan CE, Chetcuti A., Schofield PR., *Coregulation of genes in the mouse brain following treatment with clozapine, haloperidol, or olanzapine implicates altered potassium channel subunit expression in the mechanism of antipsychotic drug action.*. Psychiatr Genet., 2008. **18**(5): p. 226-239.
293. Stonehouse, A. and Jones FS, *Bromocriptine and clozapine regulate dopamine 2 receptor gene expression in the mouse striatum* Journal of Molecular Neuroscience, 2004. **25**(1): p. 29-36.
294. Lovestone S, Killick R., Di Forti M, Murray R, *Schizophrenia as a GSK-3 dysregulation disorder.* Trends in Neurosciences, 2007. **30**(4): p. 142-149.
295. Maynard TM, Sikich L., Lieberman JA, LaMantia AS *Neural Development, Cell-Cell Signaling, and the "Two-Hit" Hypothesis of Schizophrenia* Schizophrenia Bulletin, 2001. **27**(3).
296. Marley A, von Zastrow M., *DISC1 Regulates Primary Cilia That Display Specific Dopamine Receptors.* PLoS ONE, 2010. **5**(5): p. e10902. doi:10.1371/journal.pone.0010902.
297. de la Fuente A, *From 'differential expression' to 'differential networking' - identification of dysfunctional regulatory networks in diseases.* Trends in Genetics, 2010. **26**: p. 326-333.
298. Bilousova TV, Dansie L., Ngo M, Aye J, Charles JR, Ethell DW, Ethell IM, , *Minocycline promotes dendritic spine maturation and improves behavioural performance in the fragile X mouse model* J Med Genet 2009. **46**: p. 94–102.
299. Miyaoka T, *Clinical potential of minocycline for schizophrenia.* . CNS Neurol Disord Drug Targets, 2008. **7**(4): p. 376-381.